VOL. XVIII AUGUST, 1934 No. 4

THE JOURNAL OF APPLIED PSYCHOLOGY

JAMES P. PORTER, *Editor*
OHIO UNIVERSITY ATHENS, OHIO
With a Board of Cooperating Editors

SOME INTERESTING CONTRIBUTIONS IN THIS ISSUE:

- EDWARD K. STRONG, JR. Aptitudes Versus Attitudes in Vocational Guidance
- ROBERTA W. BROWN. The Touch Method of Learning Piano Music
- F. K. BERRIEN. Psychological Differences in Psychopathic and Post-Encephalitic Children
- SAMUEL GOLDBERG. A Clinical Study of K., 196 IQ
- FRED BROWN. A Psychoneurotic Inventory for Children between Nine and Fourteen Years of Age
- EDNA WILLIS MCELWEE. Profile Drawings of Normal and Subnormal Children

PUBLISHED BI-MONTHLY BY THE EDITOR
OHIO UNIVERSITY, ATHENS, OHIO
PRINTED BY R. R. DONNELLEY & SONS COMPANY, CRAWFORDSVILLE, INDIANA

Entered as second-class matter at the post office at Athens, Ohio, under the act of March 3, 1879. Additional entry as second-class matter at the post office, Crawfordsville, Indiana.
Copyright 1934 by James P. Porter

The Journal of Applied Psychology

CO-OPERATING EDITORS

W. F. Book, Indiana University, W. V. Bingham, Personnel Research Federation, New York; M. E. Broom, State Teachers College, San Diego, Calif.; Robert H. Gault, Northwestern University; William Healy, Baker Foundation, Juvenile Court, Boston; H. L. Hollingworth, Columbia University; J. E. W. Wallin, Atlantic University; Herbert S. Langfeld, Princeton University; James H. Leuba, Bryn Mawr College; Henry M. McComas, Johns Hopkins University; Walter D. Scott, Northwestern University; C. E. Seashore, University of Iowa; Daniel Starch, New York; Edward K. Strong, Jr., Leland Stanford Jr. University; L. M. Terman, Leland Stanford Jr. University; E. L. Thorndike, Columbia University; Joseph Jastrow, New School of Social Research, New York; G. M. Whipple, Danvers, Mass.; Helen Thompson Woolley, Teachers College, Columbia University.

The scope of The Journal of Applied Psychology will include the following: (a) The application of psychology to vocational activities, such as law, art, public speaking, industrial and commercial work, and problems of business appeals. (b) Studies of individual mentalities, such as types of character, special talents, genius, and individual differences, including the problems of mental diagnosis and vocational prognosis. (c) The influence of general environment conditions, such as climate, weather, humidity, temperature; also such conditions as nutrition, fatigue, etc. (d) The psychology of everyday activities, such as reading, writing, speaking, singing, playing games or playing musical instruments, sports, etc.

Communications, manuscripts, reviews, summaries, notices, books and offprints for review should be addressed to Dr. James P. Porter, Ohio University, Athens, Ohio. Contributors are requested to follow the conventional method of citation employed by this Journal; in every case the year, volume and pages should be specified.

The Journal of Applied Psychology is published bi-monthly. The subscription price is $5.00 a year. Single numbers $1.00 to $2.00, depending upon the number of pages. Remittances and business communications should be addressed to the Editor, James P. Porter, Ohio University, Athens, Ohio.

Vol. XVIII AUGUST, 1934 No. 4

THE JOURNAL OF APPLIED PSYCHOLOGY

Edited by

James P. Porter, Ohio University
Athens, Ohio

TABLE OF CONTENTS

EDWARD K. STRONG. Aptitudes Versus Attitudes in Vocational Guidance 501

ROBERTA W. BROWN. A Preliminary Study of the Touch Method of Learning Piano Music 516

K. H. R. EDWARDS. Note on the Significance of Correlation Coefficients from Different Samples 528

F. K. BERRIEN. Psychological Differences in Psychopathic and Post Encephalitic Children 536

SAMUEL GOLDBERG. A Clinical Study of K., 196 IQ 550

M. E. BROOM. A Note on Silent Reading Comprehension and Success in Academic Achievement in a State Teachers College 561

FRED BROWN. A Psychoneurotic Inventory for Children between Nine and Fourteen Years of Age 566

GRACE H. KENT. Modification of the Kohs Block Design Test 578

EDNA WILLIS MCELWEE. Profile Drawings of Normal and Subnormal Children 599

F. H. FINCH. On the Standardization of the Kuhlmann Revision of the Binet Scale 604

News and Notes 607

Book Reviews 608

New Books and Pamphlets 610

APTITUDES VERSUS ATTITUDES IN VOCATIONAL GUIDANCE[1]

EDWARD K. STRONG, JR.

Stanford University

VOCATIONAL guidance is concerned with the proper counselling of young people to the end that they may enter the vocation for which they are best fitted. The ultimate objective is greater efficiency, presumably of benefit to both individual and society in general. A secondary objective is greater happiness of the individual and consequently better social morale. This is secondary in interest, for although it has usually been maintained as an objective, it has surprisingly often been ignored in experimental work. Recently it has become thoroughly realized that preparation for an occupation involves the choice of an educational program and hence educational guidance is a necessary component of vocational guidance. Furthermore, unless there is adjustment to the total environment there will in all probability be poor adjustment to the purely occupational environment. Consequently the vocational guidance movement emphasizes the integration of the individual and his environment.

Such a conception is so broad that the scope of vocational guidance becomes almost equivalent to that of psychology on the one hand and to that of our economic environment on the other hand. It is the manner in which the common content is utilized, however, rather than the nature of the content which differentiates vocational guidance from general psychology.

It is customary to view the selection of a vocation as dependent, on the one hand, upon an understanding of one's capabilities and on the other hand, upon what capabilities are required in the various occupations. Although there are many devices which are more or less useful in helping one analyze himself and the

[1]Presidential address read before the Western Psychological Association, University of Southern California, June 16, 1933.

various vocations, yet after all the chief reliance must still be placed today upon the results of actual experience. If we all lived as long as Methuselah, it would be possible to try out a great many occupations and finally decide upon the one for which we were best fitted. As it is, we can spend only a short time trying out a few probabilities and then must settle down regardless of whether the best possibility has been found or not.

What is needed is a substitute for real experience. If instead of spending several months each working as a carpenter apprentice, a filling station attendant, a helper in a garage, etc., a young man could take a test and learn from it what his reactions would be to such activities he would be saved all the months so engaged. And if he took enough such tests he would eventually have a good idea of what kinds of occupations he was fitted for and what kinds he would dislike. If the tests could be broadened out so that they would test not mere simple jobs but occupational careers the gain would be much greater. In such a case the boy might learn, for example, that his best usefulness would lie in becoming an auto mechanic, eventually graduating into foreman and then owner. Knowing all this would help immeasurably in working as an apprentice at washing cars and all the other disagreeable tasks around a garage.

This is an idealistic program far above our powers to achieve today. But because the resulting gain in happiness and efficiency would be tremendous, the whole program intrigues us greatly.

What is the real experience which one encounters in working at any job and upon the basis of which one decides whether he likes his work or not? The aptitude-testing program more or less assumes that this real experience is largely limited to performing certain movements involving of course certain mental activities incident to coordinating eye and hand and the like. This program assumes that if the individual has the aptitudes required on the job then that job is the one for him. The program has the genuine merit of attempting to differentiate those who can from those who cannot succeed as far as the specific aptitudes are concerned. On the other hand, the vocational interest program concentrates upon whether or not the individual has the

interests that characterize the men to be found in the occupation. The assumption is that if a young man has the interests of lawyers, for example, then he will like to do what lawyers do and he will dislike what lawyers do not have to do. The results so far obtained warrant the further assumption that there must be at least a fair correlation between having the interests characteristic of men who perform the work successfully and possession of the aptitudes necessary to do the work. We are not prepared as yet to say, however, how far this holds true.

Leaving aside for the moment these testing programs and their assumptions, let us consider what is the real experience involved in a job. Mere capacity to perform the various operations constituting a job is not the only consideration in determining fitness. We may know fairly well from careful time and motion study what the man does as he tends his machine, but these movements repeated over and over represent only a portion of the total. The reveries of such a worker have recently been recognized as an important element. In addition there are a host of relationships between the worker and his fellows, his foreman, and the total environment that play a more or less important rôle. Some men hold a job because they like it, some do so only because it enables them to be with a chum, or to hobnob with their cronies after working hours. A timid, sensitive man may have all the capacities required to perform a job successfully, yet if that job necessitates associating with rough characters this workman will be unhappy and will most likely quit for something else at the first opportunity. It is an extremely complex interplay of forces which decides acceptance or rejection of a job, which determines furthermore how hard the man works at the job.

The real job experience which we are discussing is not a mere description of the job and the working environment about the job, no matter how detailed that description is. The job experience is the resultant of a given individual living in the job environment. No two men will have the same job experience even though they work side by side and perform exactly the same operations.

When we attempt to set up vocational guidance on the basis

of analyzing men on the one hand and jobs on the other hand, and then putting the two together in the best way possible, we have three exceedingly difficult tasks to perform. Of the three, the third is the hardest of all for there are many factors involved which are not to be viewed as so many blocks to be dropped into the holes of a form board. Instead of such an all too simple conception of fitting a man into a job we must realize that possession of nearly all the necessary traits may not be sufficient for success and the possession of a very few may be sufficient for success. Thus a capable employee may not be continued in his work for the single reason that he is not respectful to his superiors and another employee with many failings but capable of doing very accurate work may be retained.

Consider now some of the difficulties involved in developing aptitude tests. We start with the job experience, what the employee does, thinks, and feels while living in that working environment. This job experience is then analyzed into its parts. As no one can fully describe a job experience it is of course impossible to suppose that any one today can really analyze a job experience into all of its components. We are not even in the position to know how good an analysis we have made. Next we determine the best possible "measuring sticks" for each of the components of the job experience. In a few cases we can employ excellent quantitative measures, in some other cases we can find nothing better than rough qualitative measures; the remaining components we must ignore for we have no way of measuring them. It is those components of the entire job experience for which we have quantitative and qualitative "measuring sticks" which constitute the criterion. Let us note that this criterion cannot measure the complete job experience for it represents at best only a few of all the components of the job and furthermore it represents these few only to the extent that our methods of measuring are adequate measures. Finally, we develop tests which will correlate satisfactorily with the criterion. No one has developed a test in this connection which has perfect reliability or anything like perfect validity. Even if a test, or better still, a battery of tests, correlated perfectly with the criterion we might have a test which would be rather ineffective since the criterion

itself might be a quite inadequate expression of the total job experience.

Suppose that we concentrate upon the simpler task of developing a battery of tests which will correlate very highly with production records. One has only to read Mathewson's *Restriction of Output Among Unorganized Workers* to realize that very few if any employees are working up to their capacity. Instead of output representing capacity, it represents what employees believe to be sufficient to hold their jobs, to guard against future wage cuts, to be looked upon with favor by their associates, etc. It is very doubtful if a perfect aptitude test designed to measure capacity to do a particular job would correlate above .4 or .5 with actual production records, because those records are such inadequate expressions of real capacity. We have proceeded too much on the assumption that the output records in industry are comparable to the records of athletes on the track. Possibly the latter are the only output records that express maximum capacity.

Output records in industry reflect also many other factors besides those of the individual's capacity and willingness to work. In many cases output is determined by the capacity of the machine he operates. In a very real sense output in such a case does not represent skill in operating the machine for the necessary skill may be insignificant. Output of any employee is a measure also of managerial efficiency. Maximum output is possible only if raw materials are always at hand, finished materials promptly removed, good tools supplied, electric power always turned on, etc. It is therefore very questionable whether aptitude tests measuring capacity can be standardized against production records with any great degree of effectiveness.

In developing educational tests great care is exercised that such factors as age, intelligence, education, and the like shall be equalized. Because of the scarcity of available data useful for criteria, it has not usually been possible to take these factors into account in standardizing aptitude tests. The "validity" of the test represents then the relationship between test scores and output records of the particular men considered. Too often when such tests have been used in another connection they have given disappointing results.

There is still one other consideration that should be mentioned in this connection. When vocational tests are standardized on the basis of employee performance there are no available records from employees who have failed or who have quit the job for one reason or another. Furthermore, there are no records from men who applied for work and would have been successful if hired, but were not hired for one reason or another. Consequently the standardization of the test reflects the idiosyncrasies of the employment manager to an unknown degree and is limited to records of successful men. It is one thing to know that successful men score so and so but until we know that unsuccessful men score otherwise we cannot be sure that the test will really differentiate between the two groups.

The only way to offset these difficulties of sampling is to test large numbers of individuals and then follow up their careers and see what happens. To do this adequately requires facilities of time and money far beyond the resources of a college professor of psychology.

In one sense what has been said is a fair criticism of the aptitude-test program; in another sense it is not so fair. So far we have considered the problem of vocational guidance testing from the standpoint of the pure scientist. We have pointed out how inadequately the best available criteria represent the total job experience, how inadequately production records, the only criteria we have, really represent capacity to do the work, and we can quote Hull[1] to the effect that "the practical range of forecasting efficiency of modern psychological tests occupies the narrow zone roughly between 10 and 30 per cent" of forecasting efficiency. Our comments merely emphasize the fact that there is here a vast expanse of unknown land awaiting exploration. The explorer will have to be equipped with first-class psychological training and will have to dwell in this land of industrial activity if he is to come to know it intimately.

It has been my observation that when students read Hull's statement that 30 per cent efficiency is about the limit to aptitude testing they tend to throw up their hands and say, "What's the use?" Is this a fair and appropriate reaction?

[1] C. L. Hull, *Aptitude Testing*, 1928, p. 275.

Pure and applied scientists approach a question of this sort from two different points of view. The former seeks perfection; he longs for correlations of unity. Moreover, to a very large degree he selects the problems he would solve and he very nimbly dodges those that do not look promising. The applied scientist, on the other hand, must attempt the solution of problems which are thrust upon him and, moreover, he must solve them some way or other and usually within a short interval of time. Is it any wonder then that any contribution that bears upon the problem is highly prized by him? Is it any wonder that the pure scientist frequently views the solutions of the applied scientist with sarcasm and that the latter feels hurt that his contribution is not appreciated?

The pure scientist is often very inconsistent in his attitude, for after criticising the low efficiency or reliability of the applied scientist he goes ahead and uses equally unreliable measures in many other connections. Championships in golf, for example, are established on the basis of 36 holes of play. The correlation between the first 18 holes and the second 18 holes in such contests is usually only about .30, so Professor Terman tells me. No psychologist would accept the results of a test with reliability of .30 to .40, yet he and all the world accept such a test in the field of golf. One of my friends has calculated the reliability of diagnosis in a certain hospital to be about .40. Here matters of life and death depend upon medical procedures whose reliability is low. We accept them because they are the best we have.

From the point of view of necessity, aptitude testing with efficiency of only 30 per cent above chance may be a very genuine contribution. When a boy asks a counsellor whether he should enter engineering or medicine the chance is at least 50-50 of getting the right answer. But if the boy should really not enter either but some occupation that neither he nor his counsellor is considering, then the chances are very great that an incorrect solution will be reached. Undoubtedly in some cases the vocational counsellor will give genuine help, but in many cases he will not really be able to help and in some cases he must unquestionably be wrong. Consequently any procedure with as low

as 10 to 30 per cent efficiency represents a real gain in this field of social engineering.

The development of tests which will correlate with a given criterion can be left to psychologists. But the analysis of job experiences into their component parts and the development of methods of quantitatively measuring these components must be delegated to psychotechnologists who live within the industrial environment and devote all their energies to the task. Psychologists must come to appreciate that the development of criteria which will really represent job experiences is a far more difficult task than the development of tests which will correlate highly with the criteria.

Some may ask at this point whether there is need for such extensive research, whether vocational guidance is needed. They may point out that today a man is lucky to have any job and that the emphasis will remain upon getting a job rather than upon which one of many jobs to take. The obvious answer is that if there has been any justification in giving vocational guidance to young people who were able freely to try out a number of occupations before finally settling down, there is the far greater need for such guidance when this trial and error process of finding oneself is largely eliminated. Once a young man knows the goal he would reach, he is in a position to know what he is looking for and so most apt to find it. It is the absence of a goal that leads to such desultory attempts to get ahead as we see exemplified on every hand.

There is real evidence also that industry will need vocational selection tests far more in the future than in the past. The development of machinery has resulted in the breaking down of skilled trades into simple routine tasks that large numbers of men could perform with little training. To a large degree selective tests have not been needed here, for a tryout on the job was sufficient. The introduction of still more machinery into our industrial life will break down still further some skilled and semi-skilled occupations into routine operations. At the same time, however, there is evident a counter tendency toward the building up of more complex jobs out of simpler ones. Instead of five men, for example, tending five machines in a series, one

man may come to tend one machine which does all the work of the original five. The greater investment in the new machine and the greater need for care in operating it will make it all the more important that a properly selected man be placed in charge. Feeding material into the new machine may not require any more skill than in the case of any one of the five old machines which have been replaced, but there will be far more opportunity for a skilled operator to watch the operations of his machine and prevent breakdowns or delays. It is the machine that determines speed of operation and the quantity and quality of production, but it is the operator who by knowledge and skill avoids and quickly eliminates errors and interruptions. Aptitude tests measuring speed and precision of feeding material into the machine may very likely have to be largely replaced by tests measuring quite different capacities.

Although this new trend in machine operation is already present we cannot, of course, foretell how rapidly it will advance. It may be that society will demand employment for all regardless of efficiency of production or it may be cheaper for many years to come to use common labor at low wages than to invest large sums in new machinery.

Let us turn now from this discussion of aptitude testing and consider attitude testing. In the former we have been concerned with how well the worker could execute the various performances constituting the job; in the latter we are concerned with his reaction to the job as a whole, that is, with his reaction to what performances he is called upon to execute and to the entire job environment.

The discussion of attitude testing may best be introduced by recounting certain findings of the Western Electric Company.[1] They are so significant that no apologies need be made for bringing them again before those of you who have already familiarized yourselves with them.

---

[1]G. A. Pennock, "Industrial Research at Hawthorn"; M. L. Putnam, "Improving Employee Relations," and E. Mayo, "Changing Methods in Industry," *Personnel Journal*, Vol. VIII, No. 5. February 1930, pp. 296-332. E. Mayo, "The Human Effect of Mechanization, *Proceedings of the 42nd Annual Meeting of the American Economic Association*, Vol. XX, No. 1, March 1930.

Briefly, the production of five women operators was carefully recorded as they worked in a small room away from their department. During the first five weeks they continued working under the same conditions as formerly in the large room. Following this period, the test group was taken through twelve periods of varying lengths. In each period some change in working conditions was introduced, as for example, two five-minute rest periods, two ten-minute rest periods, six five-minute rest periods, lunch supplied; work stopped at four-thirty p.m., work stopped at four p.m.; Saturday morning off. Commencing with period twelve, these favorable working conditions were done away with, so that period twelve was identical with period three. Subsequent periods included certain favorable working conditions but not so many as in period eleven. The significant finding is that production increased steadily week by week and kept on increasing after the favorable working conditions were removed. The explanation for the 33 to 50 per cent increase in production must consequently be sought for in some factor or factors aside from improved working conditions, such as rest periods, shorter working day, free lunches, and the like, for when these were removed there was no decline in the steadily improving production record.

It is important to note in passing that the results of many extensive investigations regarding working conditions are challenged by this single study at Chicago. More favorable working conditions have been substituted in such studies for less favorable conditions and the result noted in terms of production. But such investigations did not include the further step of replacing the more favorable with the less favorable working conditions. Consequently all conclusions that this or that change in working conditions increases production by so much are to be challenged, for in this Western Electric investigation the favorable working conditions were finally removed with no decrease in production. Actually there has been increased production throughout the entire period.

What explanation can then be advanced to account for this 33 to 50 per cent increase in production? Three factors have been suggested as explanations, namely: "relief from cumulative

muscular fatigue," "change in the pay incentive," and "improved psychological attitude toward the work." The first two explanations have been disposed of on the grounds that careful investigations show no evidence of such cumulative muscular fatigue, and second, that similar results were obtained from another group of employees where no change of wage incentive was introduced. As no other explanation has been advanced, it seems necessary to conclude that the increased production is "due mainly to changes in mental attitude."

Many items of evidence are advanced by these investigators in support of their conclusion. They write:

> A relationship of confidence and friendliness has been established with these girls to such an extent that practically no supervision is required. In the absence of any drive or urge whatsoever they can be depended upon to do their best. They say they have no sensation of working faster now than under the previous conditions, and that their greatly increased production has been accomplished without any conscious effort on their part. Comment after comment from the girls indicates that they have been relieved of the nervous tension under which they previously worked. They have ceased to regard the man in charge as a "boss." Specific and individual studies which were made prove for these girls what you know about yourself—that you can work more efficiently in a contented frame of mind than you can when your mind is in a turmoil of worry, fear or discontent. You don't know exactly what it is that makes you produce more; neither do these girls. Yet they have a feeling that their increased production is in some way related to the distinctly freer, happier, and more pleasant working environment.

These investigators conclude that: "Amount of sleep has a slight but significant effect;" rest periods increase productivity; home conditions and outside influences have a distinct relationship to performance; but "the mental attitude of the operator toward the supervisor, and working and home conditions is probably the biggest single factor governing the employee's efficiency."

Further investigations should be made before the above conclusions are accepted completely. It is to be expected, however, that these conclusions will be confirmed in general since they are directly in line with the trend of personnel work in industry in which increasing emphasis has been placed upon morale in contradistinction to selection and training of employees.

If we are to accept this conclusion that the mental attitude of the employee is the biggest single factor determining efficiency, what shall be said of the whole aptitude testing program? Are we to continue to claim that future performance may be prophesied on the basis of possession of so much ability in certain traits or aptitudes? Or are we to postulate other psychological factors that can modify very profoundly an individual's output—factors that are not to be identified with simple sensory and motor aptitudes?

Possibly we should postulate at least three kinds of factors that must be taken into account in evaluating a person's efficiency. First, there would be the relatively simple sensory and motor processes. In his *Psychology of Skill,* Book distinguished between specific habits of manipulation and "general or more purely mental habits." Specifically these so-called mental habits had to do with short circuiting, meeting difficulties, acquiring and maintaining a favorable attitude, keeping attention focused, economizing effort and the like. We are not concerned here with these combining factors except to mention them in passing. Third, there would be a still higher hierarchial level of the sort one has in mind when he uses the words purpose, determination, possession of a goal, life career motive.

Just as Book postulated control of the combining of simple habits into complex ones in terms of mental habits, so control of all processes can be postulated in terms of a purpose or goal. The possession of a motive necessarily directs activities in a given direction and thereby eliminates wasteful actions. Similarly it is entirely possible to conceive of greater efficiency resulting from the operation of a purpose on the third level or a combining process on the second level, for in terms of these would come the selection of appropriate simple processes to achieve the final goal. In this way each individual tends to make use of the

best assortment of simple processes which combined will give the desired result. We are all familiar with the surprising success of individuals who have severe physical handicaps but nevertheless get the task done in some original manner. The key to their success is not the efficiency of certain simple aptitudes usually employed in the task and which they may not even possess but rather their ability to combine entirely different aptitudes.

Every individual has an enormous amount of potential energy stored up in his body. In climbing out of the Grand Canyon one lifts his body 5,000 feet. If that body weighs 180 pounds then he does the equivalent of lifting 450 tons of coal one foot! Although people undoubtedly differ greatly as to the amount of stored-up energy in their bodies, yet for our present purpose it is beside the point to talk of people varying with respect to energy. All healthy people have more than enough energy for ordinary purposes, but they differ greatly as to how much they liberate. It is this faculty of expending available energy which is the most significant aspect of our third hierarchial level. As Dodge[1] has expressed it,

> inner reinforcement by increased motivation is, I believe, at least one factor in the underlying psychophysics of James' "reservoirs of power" which may be quite as significant for psychology as the action of adrenalin to which Cannon has introduced us.

Worry, suspicion, and fear may be thought of as controlling functions which inhibit the expenditure of energy. We all realize that no one who dislikes, or is afraid of, his superior can do his best. The elimination of worry and fear has the effect of very gradually and unconsciously "releasing the brakes" so that the individual releases more energy for the work in hand, or, what is seemingly a truer picture, he does the same work with less and less energy for he is no longer working against inhibiting brakes. The latter is undoubtedly a truer picture for the universal testimony in industry is that increased production resulting from improved morale is accompanied by the feeling of less fatigue.

[1] R. Dodge, *Conditions and Consequences of Human Variability*, 1931, p. 26.

If we can obtain greater production by improving morale than we can get in any other way, which is the conclusion to be drawn from the Western Electric experiment, then it seems very questionable if any measurement of simple aptitudes will ever give us satisfactory correlations with production. In any complete analysis of an individual's capacity aptitudes should be taken into account. But it is far more likely that an analysis of his purposes, interests, goals, will afford a better basis for prophesying what the individual will accomplish.

Interests have a certain relation to capacity although the exact degree has not been determined. Each occupational group, so far considered, has a characteristic pattern of likes and dislikes. Individuals whose interests do not coincide with this pattern are seldom found in the occupation; individuals whose interests coincide somewhat with the pattern are sometimes found in the occupation, but they are not on the average so successful as are those whose interests are in complete agreement with the occupational interest pattern.

Off hand, one might suppose that the interest pattern peculiar to men in a given occupation is the resultant of the experiences to be gained from working in that occupational environment. Unquestionably there is truth in this hypothesis but it does not explain all the facts of the case. For there is clear evidence that the possession of an occupational interest pattern may occur prior to any actual experience in the occupation. Many college and even high school students rate high in occupational interests they have never studied or even considered as a life career. Conversely, neither specialization in college for four years nor actual experience for two or three years in the occupation will cause many individuals to develop the interests peculiar to that occupation.

The recent report of Carter[1] makes clear that there is some innate basis for interests since the correlation between occupational interests of fraternal twins is .28 and for identical twins it is .50. No one would uphold the idea that interests are themselves innate in any sense. But it is quite conceivable that interests develop out of early experiences and that success or failure in

[1] H. D. Carter, "Twin Similarities in Occupational Interests," *Journal of Educational Psychology*, Dec. 1932.

such experiences is conditioned by the abilities one possesses. What one likes to do is thus an expression to some degree at least of what one can do. In some such way as this we must explain the fact that interests are to some degree measures of ability.

It may well be that an adequate measure of a man's interests will prove the best approach to an understanding of him. For it will depend not only upon his aptitudes, but also upon the attitudes, purposes and goals which are such important indications of probable action and achievement.

What has been presented here has not been for the purpose of disparaging aptitude tests. We shall never thoroughly comprehend human behavior until we understand man's aptitudes and their relations to one another.

This paper has attempted to point out the exceedingly complicated problems confronting psychologists in developing a science of vocational guidance; and to emphasize that such factors as morale, interest, and purpose are among the most important, if not the most important, of all considerations. We must learn how to measure such factors if we are ever to attempt to advise and guide. More important still, we must learn how to develop and strengthen proper motives, under the inspiration of which the individual will be energized to his maximum performance.

A PRELIMINARY STUDY OF THE TOUCH METHOD OF LEARNING PIANO MUSIC

ROBERTA W. BROWN[1]

Chicago, Illinois

IN LEARNING to read piano music where eye movements are not restricted it seems likely that connections are formed between the sight of the position of notes on the staff and the sight of the position of the fingers and keys on the keyboard to the hindrance of other connections. It might be expected that continual eye movements from score to hands and from hands back to score would incur loss of time and energy.

There is a wide variability of steadiness of eye movements, particularly vertical eye movements, in professional pianists when sight reading. It is assumed that, other things being equal, the steadiness of the eyes on the score is in positive correlation to the ease and speed of learning.

The object of this study[2] was to find out the relation between two ways of learning, with special reference to vertical eye movements; one way was the usual free way, i.e., the children could look at the keyboard and their hands at will; the other way was with restricted eye movements, i.e., the children were prevented

---

[1] The writer offers thanks to Professor Arthur Bills, Department of Psychology, University of Chicago, for reading the manuscript and for making many valuable criticisms.

Thanks are also tendered to Professor Frank Freeman, Department of Education, University of Chicago, for his interest and encouragement.

To the ten pupils who joined with the writer in watching "the road wind up-hill all the way," gratitude is expressed.

Copies of the scores used may be had from the writer at cost.

[2] Two other problems of the series "Efficiency in Piano Learning" appeared in the *Journal of Experimental Psychology*, June, 1928, and June, 1933.

from looking at their hands or the keyboard by a shield which was attached to the knobs on either side of the piano just above the keyboard and then tied around the neck of the child.

A preliminary study was conducted in the summer of 1932 with children from the Chicago and Jewish Orphanages, but no conclusive results were forthcoming. Both the number of pupils and the number of lessons were too few. However, faults in the technic were brought out and valuable experience received.

In the second experiment ten children participated. Nine received thirty-eight lessons each and one, K, who started late was able to get in only twenty-two before vacation time. However, the data from only twenty-two lessons of the partner of the odd child, M, were used.

The subjects were pupils at the Elementary School of the University of Chicago, four boys and six girls. They were all in perfect health and quite robust except one, M, who was a quiet, gentle boy and had not much vigor. The children were paired as to chronological age and school grade and one of each pair wore the shield between his or her hands throughout the learning.

Before the learning began a rough estimate of each pupil's memory for pitch was attempted by the following device. The pupil was seated with his back to the piano. Middle C was sounded once every three seconds, six times. Other five notes along with C, previously arranged in random order, were played at about the same rate. The child reported "Yes" if the note he heard was the same as that he had previously heard, and "No" if not the same. Seven series for C were given, seven for G, and seven for F. The report follows below under "Pitch and Memory Errors."

None of the children had had piano lessons before. However, the children were asked to play what they had learned "by ear." S could play "Jingle Bells," "Holy Night," "Old Kentucky Home," etc., correctly with one hand, and she attempted compositions to little rhymes which were melodious to some extent. V could play "America," "Jack o' Lantern," "Pumpkins are Gay," "Choo, Choo, Choo," and "Peter, Peter," very nicely with one

hand. St played "Red Leaves," but no melody could be recognized. Later, she volunteered a song she had composed but again no melody was present. None of the other children could play anything as far as could be ascertained.

No home practising was allowed and as the children had no scores and three had no pianos this requirement was fairly well met. However, one or two of the children, especially S, who had good memories for melody were, after a while, attempting to play the melodies at home. This was not serious for the accuracy of the results, in all probability, since by the time sufficient memorizing was done to carry the melody home, the exercise was almost finished and the child had passed on to a new one.

The scores used were composed by the writer. They consisted of very simple melodies all four measures long, with not more than sixteen notes in each clef, most of the melodies being in common time. Only two key signatures were used, C and G. Treble and bass clefs were used from the beginning. Middle C and C were used as basic notes and the notes and intervals were learned in order from those notes up and back. Seconds came first as it was assumed they were easier than thirds to play and thirds easier than fourths, etc. Only five-note melodies were used. Forty-seven exercises or scores were covered in the thirty-eight lessons.

In the first twelve scores or exercises the treble and bass melodies were in unison, two voices only being used. After that progressively difficult harmony was introduced so the children had practice early in reading two lines at once. A return to unison was often made, especially when introducing some new feature. Starting with quarter notes, whole and half notes were gradually introduced and also dotted halves. Towards the end of the series of exercises, three voices, or double notes, were used.

Sixteen recapitulation exercises were interspersed throughout the series of exercises, usually one after every three exercises. These "R" exercises aimed at recapitulating the new learning incorporated in the previous three exercises, e.g.: Exercise 20 introduced the minor third D-F with the second and fourth fingers ascending only; Exercise 21, the minor third ascending and de-

scending; in Exercise 22 the melody started on F with the fourth finger and the "R" exercise which followed aimed at including all three factors. However, because of regard for attractiveness of melody and harmony the plan was by no means strictly adhered to. Another reason for the introduction of "R" exercises was that they provided an opportunity for gaining some information about the effect of speed increments in repetitions.

All playing was done to the beat of the metronome which was kept at sixty beats to the minute for all exercises except the "R" exercises where there was an increment of twelve points on the metronome after the first trial, thirty-six points of increase in all.

The children played by the whole method continually. If they lost the place or stopped for any reason they were started again at the beginning of the faulty measure. The children played each exercise three times, stopping at the end of the exercise only long enough to decide with the teacher whether the playing was correct or not. They played the "R" exercises four times with three speed increments. The two hands were never used separately.

The teacher sat at the child's left, and for the first four lessons pointed with the second and third fingers of the left hand to each note as the child played it. Occasionally the teacher sang along with the child's playing. It was proposed to eliminate all pointing and singing after the fourth lesson, but the plan was never strictly carried out, especially regarding singing. However the children were all treated alike as far as is known with regard to what little infringement there was.

Lessons lasted from twenty to thirty minutes. The six-year-olds had twenty minute lessons, the seven-year-olds twenty-five minutes, and the eight-year-olds thirty minutes. All the children except K had three lessons weekly, thirty-eight in all.

Scoring was done by the all or none plan. "Plus" was recorded after perfect playing and "Zero" after imperfect. Errors could all be objectively determined. No attempt was made to measure artistic production. Accenting was not taught but it was noticed that some of the pupils accented slightly of their own accord. The following types of errors were observed:

TABLE I

	Name	Numbers Corresponding to Pairs	Chronological Age	Mental Age	Intelligence Quotient	School Grade	Estimation of Interest in Piano Learning	Estimation of Response to Rhythm	Estimation of Response to Melody	Estimation of Memory for Melody	Estimation of "Smoothness" of Disposition	Estimation of Determination	Pitch Memory Errors Before Learning	Pitch Memory Errors After Twenty Lessons	Usual Length of Lesson in Minutes	Number of Lessons	Total Time Spent in Learning (Min.)	Number of Correct Responses in Exercises	Number of Correct Responses in Recapitulation Exercises	Total Number of Correct Responses	Order of Efficiency	Initial Correct Responses in Exercises
Free	M	1	8-1	9-6	118	2	A	A	+	+	A	A—	8	0	30	22	577	312	168	480	1	7
	R	2	7-3	8-0	110	2	+	A	A	A	+	A	28	23	25	38	968	545	230	775	1	8
	J	3	8-6	8-6	116	3	+	A	A	+	+	+	13	7	30	38	1129	545	395	940	1	16
	S	4	6-1	9-2	151	1	+	+	+	+	A	A	15	2	20	38	748	549	238	787	1	7
	T	5	7-11	9-6	119	2	A	A	A	A	A—	A	5	3	25	38	915	518	258	776	1	7
			Av. 7-7	8-11	123							Av. 13-8					4337	2469	1189	3658		45
Restricted	K	1	8-3	12-0	146	2	A	?	?	?	A	A	1	0	30	22	523	393	84	477	2	10
	U	2	7-3	7-3	100	2	+	A	A	A	A	A	28	27	25	38	956	494	228	722	2	3
	N	3	8-3	10-3	125	3	+	+	+	A	A	+	19	5	30	38	1105	553	252	805	2	11
	V	4	6-4	7-8	122	1	+—	+	+	+	—	A or +	2	0	20	38	781	527	245	772	2	5
	St	5	7-4	10-4	140	2	+—	—	A or —	—	—	A or +	10	7	25	38	936	371	154	525	2	6
			Av. 7-6	9-6	127							Av. 12					4301	2338	1063	3401		35

A = Average.
+ = Superior.
— = Inferior.
+— = 1st S, 2nd I.

1. Wrong note.
2. Wrong finger.
3. Playing slower than the metronome.
4. Stopping for one or more beats.

At the beginning great exactness was employed in recording errors. If the child tipped the side of another note while playing the correct note an error was recorded. If the child put the finger so softly on the correct note that it did not sound, an error was recorded. This exactness seemed to be discouraging, so, as far as those two cases are concerned, the exactness was mitigated. S gained confidence in new exercises very early and had a tendency to go ahead of the metronome. This tendency was ignored.

An exercise was considered completed when it had been played correctly at the first trial on four successive days. The children knew this and also that each "plus" was to be counted in the total score. Occasionally when a child failed in the second trial although he had passed in the first, a remark such as "I'm glad this was not the first trial" was heard. An "R" exercise was considered completed when the exercises which led up to it, i.e., came between it and the previous "R" exercise, were finished. Learning curves were charted from the highest exercise passed each day at the first trial.

V and St showed some discouragement about the thirtieth lesson. They were asked if they would prefer to have the shield taken off. On an affirmative reply from both, these two children completed the learning without the shield. However, as far as could be ascertained, there was no difference in their speed of learning after the shield was removed (see Fig. 4).

RESULTS

Fig. 1 shows the learning curves made from the averages of the highest exercises passed at the first trial in each lesson by each group of pupils. Fig. 2 shows the same curve minus the data from the learning of K and M. Fig. 3 shows the learning curves for "R" exercises (all pupils). The continuous line shows the learning of the "free" children and the broken line shows the learning of the "hands-covered" children. It will be seen that the children who learned without any restriction of their eye

movements learned a little faster than the children whose eye movements were restricted. The difference between the two groups became more marked with more advanced learning. The learning curves of the individual pairs, Fig. 4, show very similar results except for one pair, K and M, where the results are reversed.

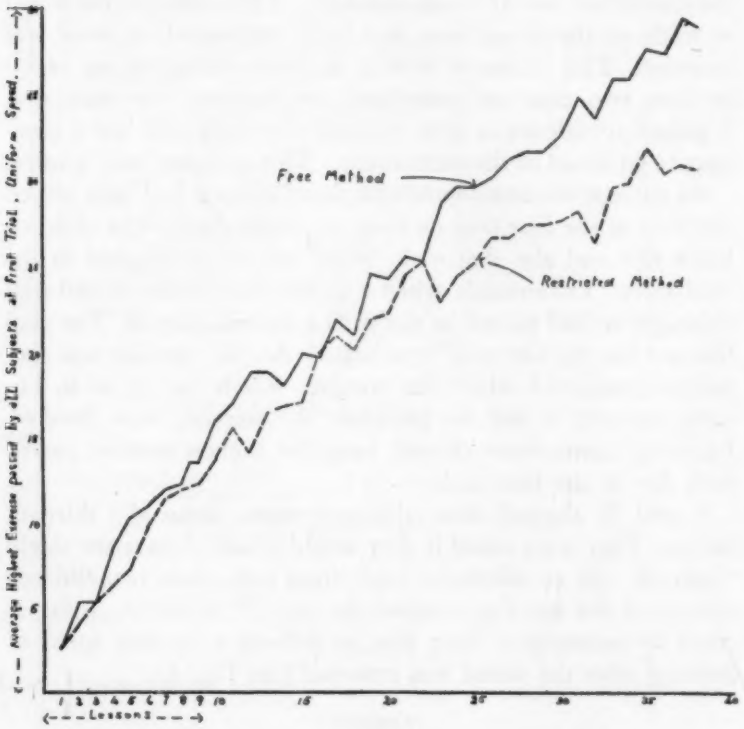

Fig. 4.

Fig. 5 shows the result of successive repetition. The mean of all the percentages of correct responses for all first trials was found for each group and also the means for the second and third trials. The results show higher means for each of the three trials of the "free" children than for the "restricted" children. Both groups of learners benefitted by a second trial. A third trial produced an increase in accuracy in the learning of the "free"

children and a slight decrease in accuracy in the learning of the "restricted" children.

An analysis of this total result was made to find out the relation between the difficulty of the task and repetition. The results of four exercises were summed in sequences of four from Exercise 4 to Exercise 23. The assumption was made that the easier

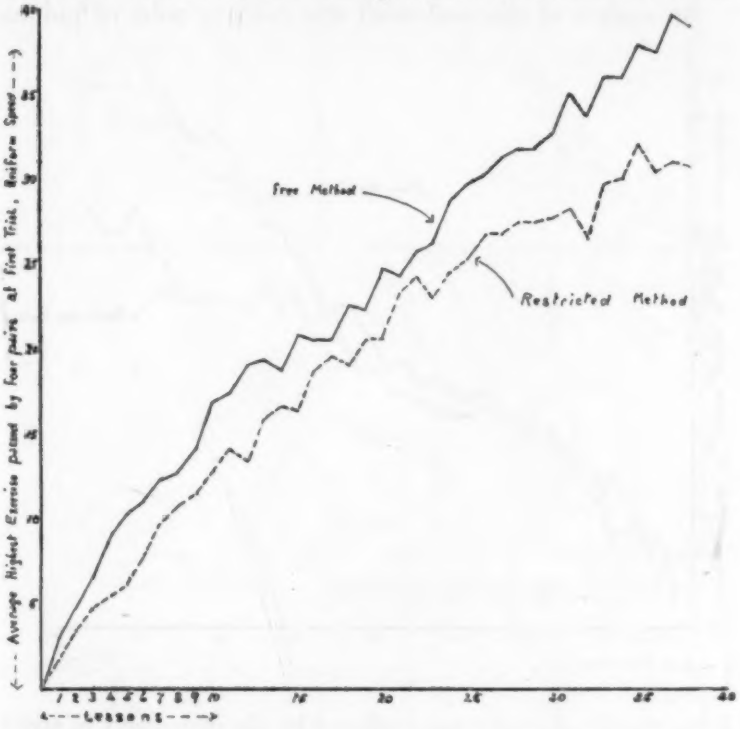

Fig. 2.

exercises came ahead of the more difficult ones. It was shown that the children of both groups showed a tendency to be less accurate when repeating the easiest exercises. The "free" children showed that the more advanced the exercises the greater the gain in accuracy through repetition. The other group did not show such constant results (see Fig. 7). Because of the dropping out of K and M the last unit represents slightly fewer records.

The difference between the two groups of learners when increments in speed were introduced was marked. In Fig. 6 it will be seen that the increased speed was detrimental to the learning of the children whose hands were covered and beneficial, as far as the first increase—20 per cent—went, to the children who were "free."

An analysis of this total result was made in order to find out

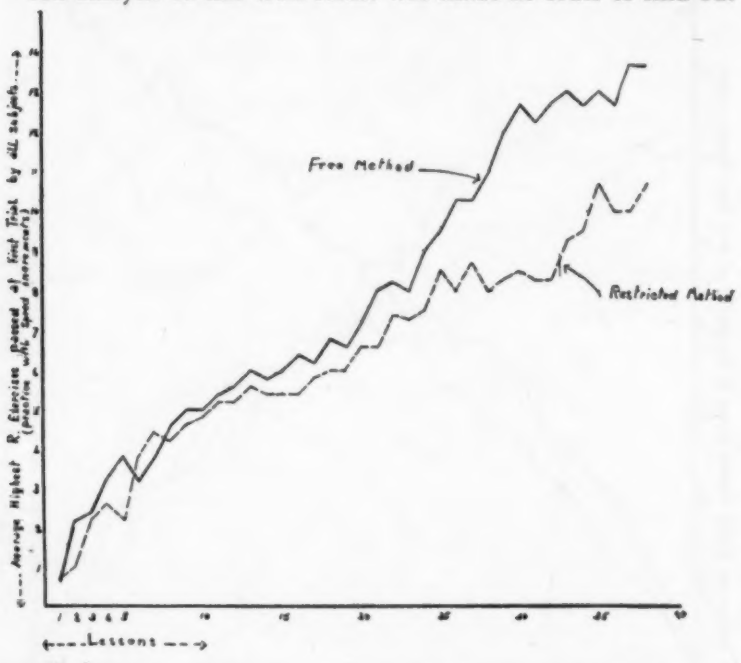

Fig. 3.

if the easier tasks were more affected by the increments in speed than the more difficult. Three groups of "R" exercises with four in each as before, i.e., from "R" Exercise 2 to "R" Exercise 13, were used, and as before the order was assumed to be an order of difficulty, the easier coming first. The results, Fig. 8, seemed to show that the easier, or first, group of "R" exercises was most affected by the increments in speed. Again, the last unit represents fewer cases because of the dropping out of K and M and because three of the other pupils had not all reached all of these harder "R" exercises.

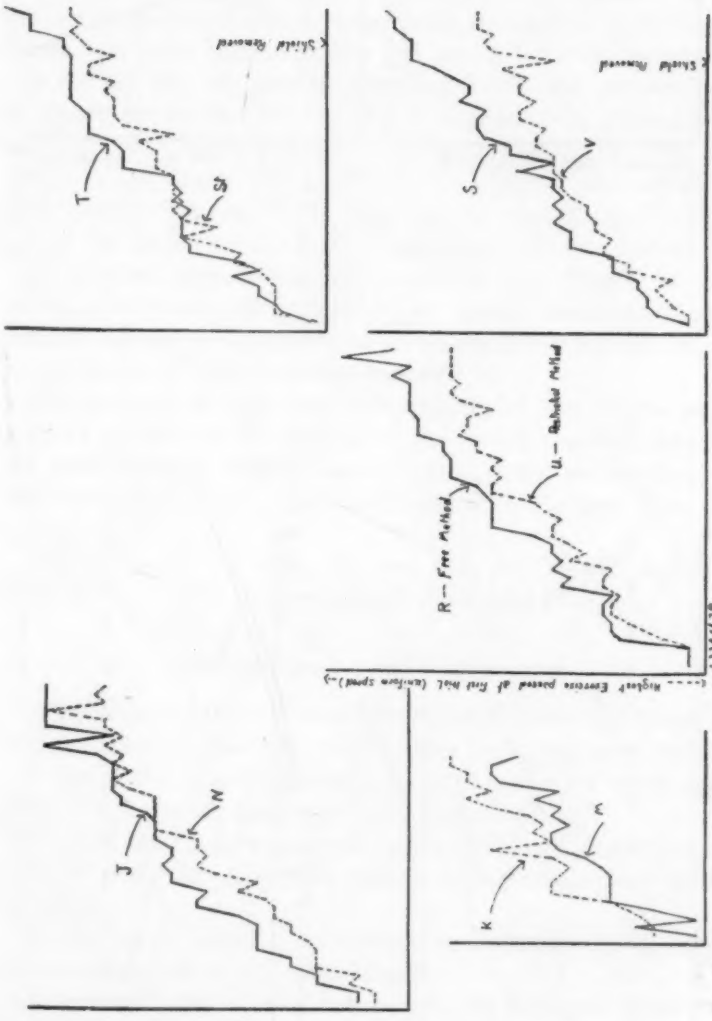

Fig 4. Showing individual learning curves. ——— lessons ——

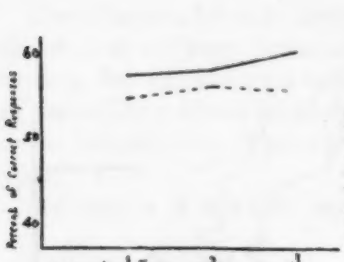

Fig 5

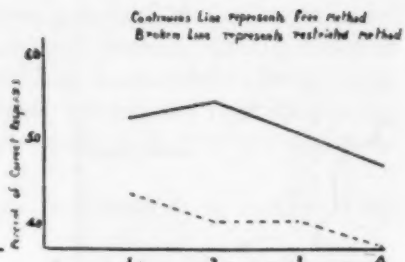

Fig 6

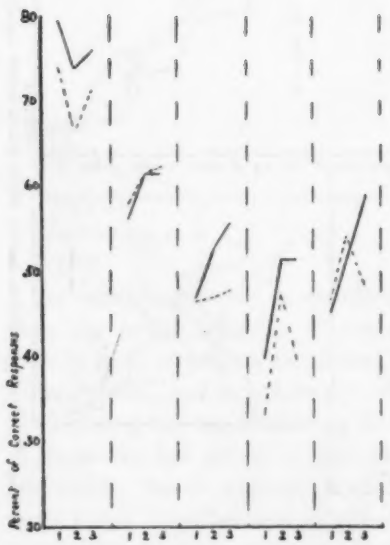

Fig. 7

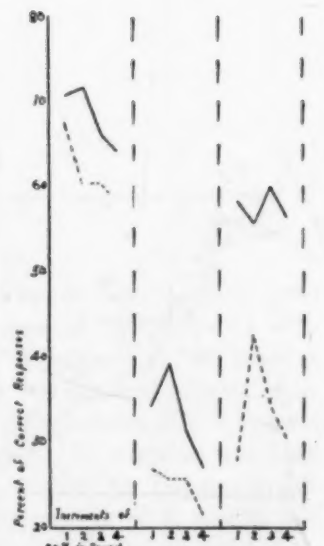

Fig. 8

CONCLUSIONS

As far as this experiment went, i.e., thirty-eight lessons, it was found that covering the hands of the learners so that they could not see the keyboard was to some extent detrimental to learning. There was some indication that age was a factor in the result.

It seemed that the greatest retarding factor was not the lack of vision, *per se,* but the amount of manipulation necessary in learning space habits through touch perception alone. For example, in learning to play a fifth, say C to G, the "covered" children usually "measured" the way up, i.e., they played C, then placed the fingers on D, E, F, in sequence and then played G. The question remains unsolved: would it take longer, on the whole, to eliminate those "guides" in "touch" learning, in case they are necessary, than it would to eliminate vertical eye movements in the ordinary learning method?

The unequal strength and independence of the fingers were more in evidence in the subjects of the "touch" method than in the usual method. Substitution of stronger for weaker fingers was eradicated slowly, less slowly in older children than in younger.

The children, as a rule, did not like the "touch" method. They found it hard. N was the only child who liked it.

QUESTIONS ARISING FROM THIS STUDY

What is the relation between general intelligence and progress in piano learning, generally, and by either of the two methods?

What is the relative importance of memory for pitch and memory for melody for progress in piano learning?

What is the relation between "smoothness" of disposition in contrast with the proverbial artistic temperament and piano learning?

If covering the hands is detrimental at the beginning of learning, is it beneficial at any other stage?

If "covered" pupils were first taught the keyboard from the staff, note by note, without melody or harmony or any other factor in music, would their learning be faster than that of ordinary learners?

NOTE ON THE SIGNIFICANCE OF CORRELATION COEFFICIENTS FROM DIFFERENT SAMPLES

K. H. R. EDWARDS

University of Durham, York, England

IT IS frequently necessary, when a group test is applied to different samples of the population, to discover if the sets of correlation coefficients obtained from the various samples are at all comparable. The test of significance usually employed is to determine the difference between r_1, the first correlation coefficient and r_2, the second coefficient, and then to calculate the probable error of this difference, given by the formula $PE_{r_1 - r_2} = \sqrt{(PE_1)^2 + (PE_2)^2}$ where PE_1 and PE_2 are the probable errors of the two correlation coefficients r_1 and r_2. If the difference between the two correlation coefficients is of the order $3 \times PE$, or exceeds this, then the difference is regarded as significant. It is assumed when this procedure is applied that the two coefficients r_1 and r_2 are independent.

For small samples, the values of the probable errors of correlation coefficients exceed those given by the formula $PE = .6745 \frac{1-r^2}{\sqrt{n}}$, this formula assuming a normal distribution. It follows that tests of significance employing probable errors calculated in this way will not be entirely trustworthy. In this paper a more accurate method of determining significance when dealing with small samples is brought to notice. An example is given showing how it is applied to test the significance of the differences between correlation coefficients obtained from three different samples.

The data were obtained as follows: The group form of the June Downey Will-Temperament Test was given to three separate populations, two groups of 50 boys and one group of 20 boys. The age of the subjects was approximately the same, lying between 15-16 years of age. The test results were marked

according to the published instructions [1] and the intercorrelations of each individual test with every other were calculated by means of the product moment formula

$$r = \frac{\Sigma (xy)}{\sqrt{\Sigma (x^2).\Sigma (y^2).}}$$

The correlation coefficients obtained are given in Table I. It will be noticed that wide variations occur in the correlations between individual tests in the three samples. Thus Will-Profile Item I (Test VII) correlates with Will-Profile-Item G (Test XII), —.234 in one sample, —.146 in another and +.468 in the third; whilst Will-Profile Item G correlates with Will-Profile Item C (Test VIII) —.145 in the first group, —.171 in the second and +.400 in the third. The values of r required to reach different levels of significance for different sized populations are given in Fisher [2] and elsewhere. For a population of 50, when P = .05 (i.e., one in twenty) the least value r should possess to be significant is ±.282. In the smaller population of 20, to reach this level of significance r should be at least ±.444. On this basis, only 6 coefficients in each half of Table I are significant in the first sample, 6 in the second sample and 5 in the third.

Consider now the differences between the samples. The method used in order to ascertain whether the differences are significant is that of Fisher.[3] The values of r, the correlation coefficients, are converted with the aid of a table [4] into corresponding values of z. This function z, introduced by Fisher has a distribution related to and is a natural extension of the x^2 distribution of Pearson [5] and the t distribution of Student.[6] The difference between the values of two z's is then found and also

[1] Downey, June E., *Manual of Directions for the Group Test*, New York, 1929.

[2] Fisher, R. A., *Statistical Methods for Research Workers*, Oliver and Boyd, Edinburgh, Table V.A. p. 176.

[3] Ibid, p. 170.

[4] Loc. cit. Table V.B.

[5] Pearson, K., *Phil. Mag. Series V*, Vol. L. p. 175.

[6] Student, Metron, Vol. V, No. 3, pages 113-120.

TABLE I

*Correlations Between Downey Individual Tests from Three Samples**

Tests	TESTS	A	B	C	D	E	F	G	H	I	J	K	L
	A		785	—061	271	—229	056	018	215	087	169	—119	—253
			600	110	112	050	—210	—080	287	—050	120	260	—050
			451	210	—361	—383	099	216	110	107	—085	674	—283
	B	785		012	171	—018	115	069	149	—151	—059	—279	—288
		600		—062	421	120	—016	100	263	—120	112	287	—034
		451		284	—344	000	153	130	139	124	—196	342	—609
	C	—061	012		—212	296	—098	—145	089	—047	333	141	100
		110	—062		121	—172	016	—171	—030	—122	350	024	025
		210	284		—564	203	081	400	028	355	116	183	—269
	D	271	171	—212		—161	274	100	044	—259	—126	—116	026
		112	421	121		124	121	033	204	—026	220	380	110
		—361	—344	—564		352	256	—329	—310	—387	000	—146	069
	E	—229	—018	296	—161		—050	—012	—043	—241	143	209	—021
		050	120	—172	124		123	086	066	090	136	070	—010
		—383	000	203	352		—199	099	062	207	—029	—565	244
	F	056	115	—098	274	—050		—328	091	—232	—099	057	—069
		—210	—016	016	121	123		180	200	—130	—057	—013	—225
		099	153	081	256	—199		—111	—204	—158	—036	002	—231
	G	018	069	—145	100	—012	—328		—279	—234	028	—329	061
		—080	100	—171	033	086	180		109	—146	—023	133	109
		216	130	400	—329	099	—111		271	468	—324	—190	013
	H	215	149	089	044	—043	091	—279		012	—032	084	016
		287	263	—030	204	066	200	109		157	—023	—061	—123
		110	139	—028	—310	062	—204	271		218	—387	—226	155
	I	087	—151	—047	—259	—241	—232	—234	012		268	045	162
		—050	—120	—122	—026	090	—130	—146	157		103	072	026
		107	124	355	—387	207	—158	468	218		—090	100	154
	J	169	—059	333	—126	143	—099	028	—032	268		081	026
		120	112	350	220	136	—057	—023	—023	103		268	—060
		085	—196	116	000	—029	—036	—324	—387	—090		111	—099
	K	—119	—279	141	—116	209	057	—329	084	045	081		101
		260	287	024	380	070	—013	133	—061	072	268		—091
		674	342	183	—146	—565	002	—190	—226	100	111		—172
	L	—253	—288	100	026	—021	—069	061	016	162	026	101	
		—050	—034	025	110	—010	—225	109	—123	026	—060	—091	
		—283	—609	—269	069	244	—231	013	155	154	—099	—172	

*The values are given to three places of decimals. The decimal points in each case have been omitted for convenience.

the standard error of their difference. If the difference between the two z values is greater than twice the standard error of their difference, a significant difference is indicated.

For example, consider the correlations between Test A and Test B. In the first sample, the correlation is +.785, in the second +.600 and in the third +.451. Are their differences significant? Taking the first two,

	n′	r	z	n′—3
First Sample	(50)	.785	1.0582	47
Second Sample	(50)	.600	.6937	47
Difference			.3645	

The standard error of this difference is the square root of the variance of the difference, which is the sum of the reciprocals of n′—3.

$$SEz_1 = \frac{1}{n'-3} = \sigma 1$$

$$SEz_2 = \frac{1}{n'-3} = \sigma 2$$

$$SEz_1-z_2 = \sqrt{\sigma 1^2 + \sigma 2^2}$$

$$= \sqrt{\frac{1}{n'-3} + \frac{1}{n'-3}}$$

$$= \sqrt{\frac{2}{47}}$$

$$SEz_1-z_2 = .206$$

Since the difference between the two z values, namely .3645, does not exceed twice the standard error of their difference .206, the difference between the two z's, and consequently that between the two r's, is not judged to be significant.

Comparing the first and third values,

Example II

	n′	r	z	n′—3
First Sample	(50)	.785	1.0582	47
Third Sample	(20)	.451	.4860	17
			.5722	

$$SEz_1-z_2 = \sqrt{\sigma 1^2 + \sigma 2^2}$$

$$= \sqrt{\frac{1}{47} + \frac{1}{17}}$$

$$SEz_1-z_2 = .283$$

In this case, the difference between the z's is slightly higher than twice the standard error of their difference and may, quite possibly, be regarded as significant; although it is a borderline case. When r has a negative value, z also takes a negative value, the difference between the z's being determined in the usual way, due regard being paid to signs.

The whole of Table I is treated in this way. The values of r in the first sample are compared with the values of r in the second sample, those in the first compared with those in the third and, finally, those in the second compared with those in the third. The differences between the calculated values of the z's, into which the r's are transformed, are presented in Table II. Only one half of the full table is given, since the values in the lower half merely duplicate those in the upper half.

TABLE II

*Differences Between the Values of z Function in First and Second Samples**

Tests

	B	C	D	E	F	G	H	I	J	K	L
A	3645	1715	1655	2832	2738	0982	0779	1372	0500	3946	2086
B		0761	2724	1386	1315	0312	1225	0316	1737	5789	2623
C			3379	4809	1143	0277	1192	0776	0095	1179	0753
D				2911	1586	0673	1669	2391	3498	5165	0844
E					1766	1032	1151	3370	0011	1420	0310
F						5225	1115	1056	0353	0701	1648
G							4049	0855	0510	4783	0572
H								1533	0090	1464	1408
I									1684	0291	1375
J										1958	0862
K											1937

The standard error, which was calculated in the first example, is .206. In equally correlated populations there is only one chance in twenty of a difference occurring as great as or greater than twice the standard error. In Table II, out of the 66 differences here calculated, forming one half of the full table, five exceed the value $2 \times \text{SE}$. The conclusion to be drawn, therefore, is that the two samples do not correspond. Apparently they are not drawn from equally correlated populations.

* Values to four places of decimals. As in Table I, decimal points have been omitted for convenience.

Compare now the first sample with the third sample by again taking values of r, converting into values of z and finding the differences between the z values.

TABLE III

*Differences Between the Values of z Function in First and Third Samples**

Tests

	B	C	D	E	F	G	H	I	J	K	L
A	5722	2743	6559	1703	0387	1995	1070	0202	2558	9204	0328
B		2801	5311	0180	0387	0616	0100	2767	1395	6430	4107
C			4234	0994	1795	5697	1172	4181	2290	0432	3761
D				5302	0194	4420	3646	1432	1261	0305	0430
E					1517	1113	1051	4569	1729	8521	2699
F						2290	2981	0769	0633	0551	1662
G							5645	7460	3641	1491	0481
H								2096	3762	3152	1403
I									3649	0553	0082
J										0302	1253
K											2750

Out of the 66 differences calculated in this case, 7 exceed twice the standard error, which has a value of .283 (Vide Ex. 2). Again it appears that the two samples are not drawn from equally correlated populations.

The remaining comparison to be made is between the second and third samples.

TABLE IV

*Differences Between the Values of z Function Between the Second and Third Samples**

Tests

	B	C	D	E	F	G	H	I	J	K	L
A	2077	1028	4904	4535	3125	2997	1849	1574	2058	5519	2409
B		3562	8035	1206	1702	0304	1325	2451	3132	0641	6732
C			7613	3815	0652	5974	0020	4957	2385	1611	3008
D				2391	1392	3717	5315	3823	2237	5470	0414
E					3283	0081	0100	1199	1718	7101	2589
F						2935	4096	0287	0280	0150	0014
G							1596	6605	3131	3292	1053
H								0563	3852	1688	2811
I									1965	0262	1293
J										1656	0391
K											0813

The standard error being, as in the previous case .283, it will be

*Decimal points omitted.

observed that 6 of these differences in Table IV are significant. It follows that since 6 out of 66 differences exceed .2 × SE. there is no evidence available to show that the two samples are drawn from equally correlated populations.

Having now considered the differences between the correlation coefficients from three samples and having estimated the significance of the differences by using Fisher's z distribution, it appears that the three sets of data obtained by giving the June Downey Will-Temperament Group Test to three different groups are not at all comparable. Had they been so, it would have been possible, by employing the z distribution again, to obtain a representative set of values for the three sets of data. An example will illustrate the method to be used. Assume that the values +.785, +.600 and +.451 are drawn from equally correlated populations. It is required to find a representative value for the three. The three correlation coefficients are converted into z values and these are given due weight in inverse proportion to their variance. The first is thus multiplied by 47, the second by 47 and the third by 17 and the total divided by 111, which gives an average value of z. This is reconverted in r by means of a table.[1]

Example III

	r	z	$n'-3$	$(n'-3)z$
First Sample	+.785	1.0582	47	49.7344
Second "	+.600	.6937	47	32.6039
Third "	+.451	.4860	17	8.2620
			111	90.6003

$$z = \frac{(n'-3)z}{(n'-3)}$$

$$= \frac{90.6003}{111}$$

$$z = .8162$$

$$r = +.6730$$

So that the representative value for the three correlation coefficients +.785, +.600 and +.451 is +.673. It should be noted that tests of significance can be applied to average values of z obtained in this way, as in the case of individual values. But before such representative figures are calculated, it must be

[1] Fisher, Table V.B.

shown that the samples are drawn from equally correlated populations.

Two points of theoretical interest emerge. In the first place, it would seem that when a given test is applied to similar age groups under similar conditions and it is shown that the resulting sets of data are not drawn from equally correlated populations, this fact is of decided psychological interest. The data should be further analysed with a view to discovering the cause of this happening. In the second place, it would appear that the procedure of evaluating tetrads or of applying other detailed methods of statistical analysis to one set of correlation coefficients obtained from a single sample, will not yield as accurate results as the procedure of employing, as the basis of calculation, a set of representative values calculated in the manner here indicated and obtained from different samples of the population. To ensure that degree of accuracy which is so essential in the treatment of experimental results, this latter method should be adopted whenever possible.

PSYCHOLOGICAL DIFFERENCES IN PSYCHOPATHIC AND POST-ENCEPHALITIC CHILDREN

F. K. BERRIEN
Allentown (Pa.) State Hospital

THE object of the present study was to discover what measureable psychological differences exist between cases diagnosed *psychopathic personality* and those diagnosed *psychosis* with *post-encephalitic* effects.

The reasons for this study are important. First, these two types of cases are quite easily confused when viewed from the psychological aspects of the clinical picture. That is to say, both are characterized by temper tantrums, uncontrolled emotional outbursts, lying, truancy, petty thieving, lack of foresight, impulsiveness, and in some cases sex offenses. Because of these similarities it is of course necessary to have a clear picture of the characteristics which differentiate these cases in order to make the proper diagnosis.

The second reason is an outgrowth of the first. The marked personality change which usually follows immediately on the heels of the actual encephalitic attack is attributed to the damage wrought in the cerebrum by the encephalitic infection. This assumption may be questioned on logical grounds since cases are on record in which there is no apparent personality change following the attack. Nevertheless, the assumption seems valid in the cases under consideration and is accepted as such for the purposes of this paper.

On the other hand the behavior observed in cases diagnosed psychopathic personality are psychogenic. This assumption is not questioned.

Thus we have here two abnormal groups which are similar in a number of behavior traits and yet differ fundamentally in their etiology. Hence any psychological differences which may be discovered can be attributed to those differences in the genesis of the disorders.

Consequently this study may be looked upon as a contribution to our knowledge of the distinction which exists between the so-called functional and organic diseases.

PROCEDURES

After assembling a list of approximately two hundred tests of all types, thirteen separate tests were selected for the purpose of this study.

In making this selection a number of points were kept in mind.

First, it was desirable to select a battery of tests which would give as well rounded a picture of the individual as possible.

Secondly, those tests were selected which had a range suited to the age range of the group under investigation. The reasons for this are obvious.

Third, those tests were selected which had already been well standardized on a large group of normal children. This consideration was necessary in order to make the proper comparisons between the group under investigation and normal children.

Fourth, preference was given to those tests which were short, easily administered and easily scored. It is a well-known fact that children of the type being studied are frequently characterized by flightiness, extreme destructibility, and fluctuations in rapport.

Consequently this fourth consideration was very important and tended to overcome many of the usual difficulties encountered in testing abnormal children.

The tests finally selected were:

1. Stanford Revision of the Binet-Simon.
2. Porteus Maze.
3. Kent-Rosanoff Free Association.
4. Whipple-Healy Tapping.
5. Healy Puzzle A.
6. Goodenough Drawing.
7. Audio-vocal Forward Digit Memory Span.
8. Audio-vocal Backward Digit Memory Span.
9. Healy Picture Completion II.
10. Witmer Formboard.

11. Stanford Arithmetical Reasoning.
12. Stanford Arithmetical Computation.
13. Kuhlmann-Anderson Series (particularly tests No. 11, 12, 13, 16, 17, 18, 19, 20, 22, and 27).

In all thirty-seven juvenile patients were included in this study. Of these twenty-four were diagnosed psychosis with post-encephalitis effects, thirteen were psychopathic personality cases. The sexes were quite evenly divided. In the post-encephalitis group twelve were girls and twelve were boys while in the psychopathic group eight were girls and five were boys.

The groups were not quite equated on the basis of age. The mean life age of the post-encephalitic children was 14 years, 3 months with a S.D. of 1 year, 11 months; while the psychopathic children had a mean life age of 12 years, 4 months, S.D. 3 years, no months. All the children were committed to the Allentown State Hospital and with a few exceptions were all patients of the recently organized Children's Unit where special attention is given to the observation, diagnosis, and treatment of juvenile patients.

RESULTS

The Stanford-Binet

This test has been used far and wide and has been credited with being the best all around measure of "general intelligence."

An inspection of Tables I and II reveals that it correlates reliably with more tests than any other measure in the battery. (The median Kuhlmann-Anderson scores correlate reliably with only one less test.) However, when we consider the correlations of the Kuhlmann-Anderson Series we find that these latter coefficients are higher than the Binet in sixteen out of twenty-four instances. This condition tends to support the generalization that the Kuhlmann-Anderson Series is at least as good a measure of general intelligence as the Binet.

An item-by-item analysis was made of the scores on the Binet by tabulating those items which were passed above the mental age year. Such items were interpreted as being too easy for the patients or are placed too high in the age scale. A study of these tabulations leads to the conclusion that items XII-4

(mixed sentences), and XII-6 (five digits backwards) are more apt to be passed by post-encephalitic children than by psychopaths, while the reverse is true of item XII-3 (superior ball and field plan).

A similar analysis was made of items which were failed below the mental age year. There was a surprising lack of such failure among the psychopaths which precluded any comparison between the two groups. This condition seems to indicate that any scattering that may occur in psychopaths tends to be above the mental age year rather than evenly distributed above and below. Such a difference in test performance seems to be an important one since the mean range for both groups is 3.3 years.

The mean life age and Stanford-Binet age rating for both groups together with the S.D.'s for all measures were computed. The life age used in these computations was different than that noted in the previous section due to the fact that the Binet examinations were made at the time of admission whereas all the other tests were made during a five week period subsequent to admission.

The conclusion that seems warranted from a study of this table [1] is that the ratio (3:16) of the obtained difference between the life age and the Binet age rating to the standard deviation of that difference is so great for the post-encephalitic children that we can be quite sure such children will generally have mental age scores considerably below their life ages.

THE PORTEUS MAZE TEST

The Porteus Maze Tests have had a wide circulation ever since their introduction by Porteus shortly after the Binet began to be used in this country. The author states that the tests were devised to examine the individual's "planning capacity, prudence, and mental alertness in a new situation of a clinical nature." [2] It is true that the factors involved in the performance of these

[1] This table along with others referred to but not published herewith may be obtained by addressing the author.

[2] Porteus, S. D. "Guide to the Porteus Maze Test." Vineland, N. J. Training School, 1924, Pub. No. 25, p. 5.

TABLE I

Showing Rank-Difference Correlation Coefficients for Post-Encephalitic Children

Test	2	3	4	5	6	7	8	9	10	11	12	13	14
1 Age	.479	.283	.252	.453	—.157	.272	.436	.534	.457	.626	.710	.713	.724
2 S-Binet		.653	.717	.519	—.048	.817	.789	.688	.532	.652	.591	.623	.669
3 Porteus Maze			.638	.491	.330	.674	.519	.424	.525	.332	.324	.288	.676
4 Picture Completion II				.549	.313	.678	.613	.314	.728	.093	.208	.167	.685
5 Tapping					.082	.422	.065	.498	.614	.216	.313	.294	.709
6 Healy A						—.004	.225	.114	.006	.358	.525	.379	—.095
7 Drawing							.678	.607	.393	.536	.507	.548	.692
8 Digits Forward							.60	.601	.452	.521	.538	.528	.609
9 Digits Backward									.130	.629	.539	.573	.281
10 Witmer Form-board										—.185	.347	.343	.605
11 Arith. Reasoning											.897	.952	.660
12 Arith. Computation												.973	.649
13 Total Arithmetic													.628
14 Kuhlmann-Anderson													

TABLE II

Showing Rank-Difference Correlation Coefficients for Psychopathic Children

Test	2	3	4	5	6	7	8	9	10	11	12	13	14
1 Age	.869	.071	.163	.758	.764	.400	.880	.301	.869	.790	.682	.697	.906
2 S-Binet		.748	.355	.726	.671	.612	—.167	.451	.772	.712	.658	.691	.897
3 Porteus Maze			.341	.474	.642	.455	.075	.281	.649	.308	.306	.354	.525
4 Picture Completion II				.575	.409	.565	.187	.645	.119	.446	.296	.459	.554
5 Tapping					.643	.675	.058	.649	.542	.691	.654	.744	.905
6 Healy A						.277	—.098	.443	.599	.474	.238	.376	.678
7 Drawing							—.155	.551	.189	.237	.195	.397	.599
8 Digits Forward								.296	.016	.051	.037	.042	—.095
9 Digits Backward									.050	.420	.387	.549	.547
10 Witmer Form-board										.796	.783	.789	.746
11 Arithmetic Reasoning											.906	.958	.822
12 Arithmetic Computation												.973	.805
13 Total Arithmetic													.855
14 Kuhlmann-Anderson													

tests appear to be different from those employed in solving the situations presented in the Binet. Nevertheless our data indicate that the Porteus has a reliable positive relationship with the Binet in both of our experimental groups.

The rather enthusiastic claims made by the author are in part supported in our data since the correlations with age in the encephalitic group are in general lower than for the psychopathic group. This is revealed in Tables I and II.

Clinically the post-encephalitic children appear to be more impulsive and lacking in foresight and we note that Tables I and II indicate that these children are consistently more deficient than the psychopaths in this test.

KENT-ROSANOFF FREE ASSOCIATION TEST

This test is the only one in the battery to which the term "intelligence" cannot be applied. It has been widely used and needs no description. It was first introduced to the profession in 1910 when it was described in the *American Journal of Insanity*. The studies of free association have been numerous and too voluminous to be reviewed here.

One important conclusion which seems to be justified from review of these previous studies is that much depends upon the method of presentation. In our present investigation stimulus words were given as rapidly as possible. This was done in order to overcome the tendency toward distractibility on the part of the patients. It was felt that if the words were given rapidly, attention would be less likely to wander from the subject at hand. In one or two cases it was necessary to split the test and give part at one time and part at another when it became apparent that the patient was too distracted or failed to respond to the stimulus words.

In analyzing the results the classification defined by Rosanoff was used; namely, Common Reactions, Doubtful Reactions, Failures to React, Individual Reactions. To these categories was added another, Retarded Reactions, which included those responses requiring more than four seconds. In no case was a patient allowed more than eight seconds in which to respond. The Individual Reactions were divided into the sub-classes:

Normal (by appendix), Juvenile, Derivatives, Partial Dissociations, Perseverative Reactions, and Unclassified.

The following tentative generalizations seem warranted. In the common reactions the two abnormal groups approximated the two-hundred-forty-seven insane adults studied by Rosanoff.[1] The doubtful reactions were similar in frequency to the one-hundred-twenty-five normal white children, eleven to fifteen years of age; whereas the large percentage of failures to respond in our data seems to be characteristic of younger children. However, this large percentage of failures to respond in the two abnormal groups may be due to the method of presentation.

The subdivisions of the individual responses reveal marked differences from the normal groups. The normal (by appendix) reactions for the psychotic children are similar in frequency to those of normal children but the former appear to be more apt to give derivatives of a stimulus word than normals. The post-encephalitic children gave more partial dissociative and perseverative responses than the psychopaths.

In summarizing the Free Association results we can say that only slight differences appear between the two psychotic groups but marked differences appear between these psychotic groups and normals. These latter differences are most marked in the frequency of common responses.

THE TAPPING TEST

The Whipple-Healy test of tapping is the only definite psychomotor test in the battery and has had a wide use in all psychological clinics. In scoring this test the norms published by Hewes[2] were used. There is apparently a great sex difference on this test and in compiling the present data the mean age rating for each sex was first derived and then averaged in order to compare the groups.

The results published in Tables IV and V reveal that this test along with the others that follow gave the post-encephalitic chil-

[1] Rosanoff, A. J. "A Manual of Psychiatry." John Wiley & Sons, New York City, 1927, p. 550.

[2] Hewes, A. "Standardization of the Whipple-Healy Tapping Test" Jr. Ap. Psy. 1922, pp, 113-119.

dren an age rating much below their life age, whereas there was very little difference between the age rating on this test and life age for psychopaths.

KUHLMANN-ANDERSON SERIES

This series of tests has been less widely used than the Stanford-Binet but it has been proposed as a measure of general intelligence.

The standardization of these tests was made after testing over thirty thousand children in the grades and high schools since 1916. In attaining the present norms, however, three-hundred-fifty non-selected school children were used in each age.

The present scale is made up of thirty-nine tests and is a reduction from fifty-one originally used in this series. From a statistical point of view the standardization seems to be quite accurate and reveals very careful work.

During the present study twenty-seven of the thirty-nine tests were used but those mentioned in Table VI were employed most frequently. At least nine tests in the series were administered and the median score was taken as being the M.A. score for the series. In analyzing the results of this series a technique similar to that employed with the Binet was used in finding those tests on which the patients received scores above and below the median score.

Referring to Table III we note that tests 11, 12, 17, and 20 seem to show contrary tendencies in the two groups.

TABLE III

Showing Tendency of Certain Tests in the Kuhlmann-Anderson Series to Raise or Lower the Age Score

	Test no.	11	12	13	16	17	18	19	20	27
Post-Encephalitic	Freq. below median age score	7	11	7	5	5	5	1	2	2
	Freq. above median age score	10	3	9	11	8	11	8	4	8
Psychopathic	Freq. below median age score	6	5	6	2	8	6	2	6	3
	Freq. above median age score	2	6	4	8	1	6	6	3	9

Test 11 consists of a series of twelve one inch squares in which varying numbers of dots appear. The patient is required to count these dots as quickly as possible and write the number under each square. The time allowed is one minute; the score depends upon the number of correctly counted squares. This test tended to increase the scores of post-encephalitic children and decrease the score of psychopaths.

Test 12 is a directions test, consisting of a series of squares and circles. The patient is required to draw circles, underline certain squares, and draw lines from certain squares to certain circles. A series of six tasks is required in each of which no more than two directions are given. The score consists of the number of correctly performed tasks. This test tended to increase the median score of the psychopaths and to decrease the score of post-encephalitic children.

Test 17 is a series of pictures of common objects which are grouped in such a way that the patient is required to pick objects which are similar in nature. The score here depends upon the number of correct similarities chosen. This test tended to increase the score of the post-encephalitic children and decrease the scores of psychopaths.

Test 20 is a paper and pencil formboard test consisting of five trials with different forms. Only two forms are required to be placed in a larger form. The score is the number of correct trials performed. This test tended to increase the scores of post-encephalitic children and decrease the scores of psychopaths.

In passing we might point out that the correlations between the median Kuhlmann-Anderson score and age are higher in both experimental groups than for the Binet. This would indicate that the former test is more likely to give a valid age score than the Binet in cases of this type.

ALL OTHER TESTS

Tables IV and V give the results of test scores compared with life age. Table IV reveals that in all tests used the post-encephalitic children receive age scores on tests much below their life age. This difference is so great that we can be reasonably sure that no post-encephalitic child will score above his life age.

TABLE IV

Summary of Test Results in Terms of Years and Months

		Post-Encephalitic	Psychopathic
Life age	Mean	14/3	12/3
	S. D.	1/11	2/11
Porteus Maze	Mean	8/3	10/6
	S. D.	2/11	2/10
Tapping Test	Mean	10/9	11/10
	S. D.	3/6	3/4
Healy A	Mean	10/2	10/6
	S. D.	2/11	3/4
Digits Forward	Mean	10/0	10/0
	S. D.	4/0	4/0
Digits Backward	Mean	9/0	9/0
	S. D.	2/0	2/0
Witmer Form-Board	Mean	7/0	9/1
	S. D.	5/0	2/10
Picture Completion II	Mean	8/1	10/6
	S. D.	5/2	1/6
Arithmetic Reasoning	Mean	11/1	12/7
	S. D.	2/2	2/4
Arithmetic Computation	Mean	10/7	11/5
	S. D.	1/11	1/8
Total Arithmetic	Mean	10/8	11/9
	S. D.	1/10	1/9
Kuhlmann-Anderson	Mean	9/3	10/3
	S. D.	2/0	1/5

TABLE V

Showing $\frac{D}{S.\ D.\ diff.}$ *of Test Scores and Life Age*

	Post-Encephalitic	Psychopathic
Porteus Maze	8.3	1.6
Tapping Test	4.3	.41
Healy A	5.3	1.5
Digits Forward	9.04	2.7
Digits Backward	9.3	3.6
Witmer Form-board	6.7	2.8
Picture Completion II	5.5	1.9
Arithmetic Reasoning	5.4	.24
Arithmetic Computation	6.9	1.09
Total Arithmetic	6.7	.60
Kuhlmann-Anderson	8.8	2.3

This condition does not seem to exist among psychopaths generally. On only four tests is there a difference in the means for this group which might be considered reliable. These tests are the Forward and Backward Digit Memory Span, the Witmer Formboard, and the Kuhlmann-Anderson Series.

References to Table II indicate that there is a reliable correlation between age and forward digit memory span as well as a very low correlation between age and backward digit memory span. This great difference between the two correlations together with the data in Table V seem to indicate a test performance that is of great diagnostic significance.

The Witmer Formboard appears to deserve special mention since the ratio of the obtained differences in the means between life age and age rating on the test to the standard deviation of the differences approaches reliability. Bearing in mind the correlations with age shown in Table II we can say that the Witmer age scores are consistently low for psychopaths.

The Kuhlmann-Anderson median scores for the psychopaths seem also to be consistently below life age although this difference is the least reliable in the group of four we have chosen for special mention.

The results of the detailed analysis of the drawings studied by the use of the Goodenough scale have been included in a previous paper and will not be reviewed here.

Tables I and II although frequently mentioned in the preceding section warrant more careful study.

All the coefficients are positive unless otherwise indicated. The probable error for each coefficient is not given but we can say that for Table II coefficients below .55 have P.E.'s greater than one-fourth of the coefficient. Since there are more post-encephalitic children in our group we may consider coefficients in Table I which are below .47 to have P.E.'s greater than one-fourth the coefficients. A study of both tables reveals that the correlation between certain tests are significantly high in one group of cases while in the other the correlations between the same tests are either of doubtful reliability due to the large P.E. or are significantly lower. Accordingly we seem to be justified in making the following conclusions:

First, that the correlation between the following factors are more apt to be lower in post-encephalitic cases:

a. Age and Healy Puzzle A.
b. Age and Witmer Formboard.
c. Age and Forward Digit Memory Span.
d. Age and Binet.
e. Age and Kuhlmann-Anderson Series.
f. Healy Puzzle A and Witmer Formboard.
g. Healy Puzzle A and Kuhlmann-Anderson Series.
h. Healy Puzzle A and Tapping.
i. Arithmetic Reasoning and Tapping.
j. Arithmetic Reasoning and Witmer Formboard.
k. Arithmetic Computation and Witmer Formboard.
l. Total Arithmetic Score and Tapping.
m. Total Arithmetic Score and Witmer Formboard.
n. Kuhlmann-Anderson Series and Tapping.

Secondly, the following factors are more apt *not* to correlate highly in psychopathic children:

a. Binet and Picture Completion II.
b. Binet and Forward Digit Memory Span.
c. Porteus Maze and Forward Digit Memory Span.
d. Drawing and Forward Digit Memory Span.
e. Drawing and Arithmetic Computation.
f. Witmer Formboard and Picture Completion II.
g. Arithmetic Reasoning and Forward Digit Memory Span.
h. Arithmetic Computation and Forward Digit Memory Span.
i. Kuhlmann-Anderson and Forward Digit Memory Span.
j. Picture Completion II and Forward Digit Memory Span.
k. Forward Digit Memory Span and Backward Digit Memory Span.
l. Total Arithmetic Score and Forward Digit Memory Span.
m. Witmer Formboard and Forward Digit Memory Span.

It will be noted that in the foregoing section any references to these tests as measures of particular mental processes have been carefully avoided. There is no apparent need at present to call a particular test a measure of rote memory, mental control, reasoning ability, or what-not. The labeling of tests in such a manner does not seem to be particularly important for us here even though it may provide an interesting matter for conjecture and controversy. The important thing is that these children differ in their ability to solve the problems which these tests present. It is on that basis that the following conclusions are made:

SUMMARY AND CONCLUSIONS

Twenty-four children with diagnosis of psychosis with post-encephalitis and thirteen psychopathic children were studied by

means of a carefully selected battery of tests. The selected tests had been adequately standardized and for the most part have been used widely in clinical procedures. The purpose of the study was to discover in what respects these two types of cases differ. In the light of the previous discussion and the statistical data presented, the following conclusions seem to be warranted:

First, that post-encephalitic children differ greatly in test scores from psychopathic children.

Second, that post-encephalitic children rate consistently below their life age on all tests used.

Third, that certain tests in this battery correlate with each other to a greater degree in one type of case than in the other.

Fourth, that any "scatter" on the Stanford-Binet is more apt to occur above the mental age year in psychopaths.

Fifth, that the Stanford-Binet items VII-4 and XII-6 are more apt to be passed by post-encephalitic children while XII-3 is more apt to be passed by psychopaths.

Sixth, that important differences occur in the frequency of common responses on the Kent-Rosanoff Free Association Test when these psychotic children are compared with normals. However, only slight differences between post-encephalitic and psychopathic children occurred on this test.

Seventh, that tests 11, 12, 17, and 20 in the Kuhlmann-Anderson Series are probably of value in distinguishing between post-encephalitic and psychopathic children.

Eighth, that these differences may prove of diagnostic value from a psychometric standpoint.

A CLINICAL STUDY OF K., IQ 196

SAMUEL GOLDBERG

College of the City of New York

THE rarity of occurrence of an IQ of 196 in part explains the writer's great interest in K. Psychological literature at the time of the present writing contains few clinical summaries of individuals who obtain IQ's of 180 or higher. The relative infrequency of pupils of such intellectual endowment and the consequent lack of opportunity of trained clinical psychologists to observe such "specimens" have prompted the presentation of the following information regarding our prodigy.

Certain developmental, especially academic, problems will be touched upon in our consideration of K. In their relation to an extreme superiority of this type, i.e., an IQ of 196, they assume an exaggerated significance. It should be noted, however, that these conflicts exist not only for the upper 0.1 of 1 per cent of our student population, in which K. finds himself, but similarly exist for the upper 1 per cent and even here in relatively intense manner. As a matter of fact the upper 15 per cent of pupils because of their superior capabilities would profit from special attention.

It is not the present desire to enlarge on these many difficulties encountered in rearing "exceptional" or "gifted" children or to indicate the many inadequacies in the treatment of such cases of the regular school system and of its orthodox curriculum. Readers who are so interested can find many excellent considerations in the literature on superior children.

K. was born June 25, 1927. A written record of his infancy was not kept. Since he is an only child the following statements may be considered as fairly accurate. He started to walk at the age of 14 months. While it is not known when he started to talk, this being a difficult judgment to make in any case, he was talking relatively fluently at 1½ years of age. Dentition started when he was eight months old. He has had no serious illnesses.

When K. was 20 months old he knew his alphabet and within a short while after that he was able to recite it backwards. At about the same time he had a set of 72 blocks, which offered him additional opportunity for developing a well-nigh astounding feat. He could by looking at one of the figures on a given block call off from memory the other five objects on the remaining sides. This he was able to do for almost the entire set. At about 2 years of age K. knew his own name and address and what is more significant the addresses and telephone numbers of the entire family, numbering about a dozen.

It remained, however, for him to attain the mature age of 3½ years, in 1931, in order to perfect the following amazing technique. By studying a calendar he learned to tell on what particular day a certain date would fall. For example, if he were asked on what day of the week July 16 would fall he would indicate Thursday. Then he was also able to figure out that since 1932 was a leap year, July 16, 1932, would fall on a Saturday. In 1933 it would of course fall on a Sunday. As may be expected the family made much of this particular ability.

K. began to read at about 4 years of age. He was given no formal training in the beginning mechanics of reading. The only assistance he received was a suggestion that he "pronounce words by syllables." At this time he was already reading simple words easily.

K. is of Jewish origin. Mr. K. at the time of K.'s birth was 32 years of age while Mrs. K. was 35. Neither K.'s mother nor his father has had the benefit of college training. They are for the most part self-educated. Mr. K. is a proprietor of a small retail business.

Some of K.'s living relatives are well placed in the professions and in business. He has several first cousins of school age of whom it was not possible to secure measures of intellectual capacity. An examination of their age-grade placements, however, reveals evidences of precocity.

K. has always been recognized as a very prodigious youngster. Some persons have assured Mrs. K. that he is a genius. Many others, professional as well as otherwise, have felt obligated not only to call Mrs. K.'s attention to the fact that K. is exceptionally

brilliant, but have also felt obliged to offer recommendations for his care. Some of these follow: "K. is a genius but you ought to make sure not to let him burn himself out." "K. is very superior but he ought to be allowed to find himself socially." "K. is all right but you ought to take good care of him physically." These suggestions, as is known, are grounded in superstitions regarding intellectual superiority which have long been disproved by psychologists.

It is, therefore, somewhat surprising to find people still believing that superior children are "nervous," are physical wrecks, and that they tend to become normal in mentality as they grow older.

These erroneous beliefs come about in many ways. The following condition, however, has misled teachers as well as those not connected with the profession to accept these misconceptions and so it is described.

When a very superior child enters the 1A grade he is usually judged at least an average physical specimen and is usually rated superior mentally. This is so because he is being compared with children approximately his own age. Often he is immediately seized upon as a very bright youngster and is skipped. With the passing of one or two years, possibly three, the youngster, because of his brightness, reaches a class considerably above that attended by average children of his own chronological age. In one of the upper grades, the teacher of the 8A for example, receives her new class. In her new class is our bright youngster, much above his "normal" grade, through skipping. Her judgments, and these have been made, are as follows: "Look at——. He's so much smaller and weaker looking than most of the other boys in the class." Usually the teacher also feels and states that his social responses are relatively immature. From these invidious comparisons arises additional testimony for some of the previously mentioned misconceptions. Little does the teacher realize that she is comparing our youngster not with average individuals of his own chronological age but with average individuals of the class—*pupils who are 3 or 4 years older than he.* Little does she realize that it has been found that not only superior children in general, but that even " . . . children

who test above 180 IQ greatly exceed the norms for height and weight, as a group, and that excellent health is characteristic of them. Of those here described, the majority are reported as in excellent health while none is characterized as in poor health. They are not nervous, none being described as ill-balanced in emotional reactions."[1]

A condition, concomitant with that described above, usually results. The superior child in the lower grades generally secures the maximum rating in proficiency. Again, after he has been skipped several times and is in competition for marks with pupils much older than himself and occasionally with some who are also superior, his marks may fall to a "B" in work, if "A" is to represent maximum proficiency. If this results on several occasions, feeling becomes manifest that he is no longer as smart as he used to be and so an additional feeling that he is now only "normal" in general intelligence. Complete oversight of the chronological age of this superior child has resulted in an erroneous judgment of his present mentality. That this frequently occurs has been shown in many psychological investigations ever since the classical experiment of Binet.

The above considerations may seem somewhat naïve. They are included, however, as a further attempt to clarify certain elementary concepts for some naïve people who are still subject to misinformation regarding intellectual precocity. The writer is well aware of many of the other psychological issues which enter into the determination of grades and into the level of achievement to be expected of superior children.

K.'s visit to the Clinic was preceded by a note from Mrs. K., in which she indicated her belief in his superiority. Upon seeing K. for the first time, there was little if anything in his physical appearance and general demeanor which might confirm the subjective estimates of the many who had judged him superior. He seemed the usual youngster, i.e., until he started to perform. His first task, for example, was that of writing his name on a piece of paper. He returned the paper saying, "I think this is my best *signature.*" The perking of the eyes to this response

[1] L. S. Hollingworth, *Gifted Children*, Page 262, New York, MacMillan Co., 1926.

in no wise compares with the subsequent jars to complacency induced by the remarkable qualitative responses of K.

At the Clinic K. received a physical examination. He was given a series of intelligence tests, consisting of some of the form boards of the Pintner-Paterson Scale, the Stanford-Binet, the Kohs Block Design Test, the IVth grade level of the Kuhlmann-Anderson, and the Goodenough. Finally, he was also given the paragraph meaning and arithmetic reasoning tests of the New Stanford Achievement battery.

Though the entire clinical examination required two sittings in order to eliminate so far as possible the factor of fatigue, the presentation of our results does not show the manner in which the tests were spaced. It is worth noting that the Binet was given in one sitting with an occasional rest, and that the second part of the examination was completed within one and a half weeks after the first sitting.

Before proceeding to a discussion of his responses on the mental tests the findings of K.'s physical examination are presented. His height with shoes is 47.3 inches, which is 2 inches above the average for his age. His weight is 52 pounds, which is 1 pound above the average for his height and age. He is well nourished. He was also examined for conditions of the glands, heart, lungs, vision, audition, and teeth. These findings were all negative.

In order to surround the clinical examination with a game spirit, he was first given several of the form boards of the Pintner-Paterson Scale. When scored on the basis of time alone he secured an M.A. of about 7 on the Mare and Foal, an M.A. of about 8 on the Seguin, an M.A. of about 7-8 on the Two Figure, an M.A. of about 7 on the Five Figure, and an M.A. of approximately 14 on the Casuist. These ratings are in general considerably lower than those he secured in other mental tests.

The following explanation of these relatively inferior ratings is perhaps helpful. These tests are influenced in part by rapidity of motor coordination which in some measure is the result of physical growth. Since K. is being compared with norms established on the samplings of average children at higher chronological ages, it is difficult for him to obtain a superior rating. It is significant that he did show superior critical judgments,

good perception of form and that there was a minimum of trial and error placements.

After these items of the Pintner-Paterson Scale, he was given the Stanford-Binet. The orientation item on the IX year level was given him first. In response to the question, "What year is it?" he said, "1934, but if you believe in the Jewish calendar, it is 5694."

An analysis of K.'s Stanford-Binet follows:

TABLE I

Summary of K.'s Stanford-Binet Record

Yrs.	Mos.	Level	Yrs.	Mos.	Specific Items Passed
		7			
		8			
		9	Basal		
		10		6	Vocabulary, Detecting Absurdities, Reading for 8 Memories
		12		15	Ball and Field, Dissected Sentences, Five Digits Reversed, Interpretation of Pictures, Giving Similarities
		14		16	Induction, Difference between President and King, Arithmetical Reasoning, Reversing Hands of Clock
		Average) 16 Adult)		10	Problem of Enclosed Boxes, Six Digits Backward
		Superior) 18 Adult)		0	

Mental Age..... 12 years 11 months
Chronological Age..... 6 years 7 months
IQ.....196

An exact reproduction of his Binet is of course impossible. It would be difficult to include all of his responses, indicating wherein he might have scored higher or lower. For example, in the naming of words on the X year level, it is noted that he

failed this item; he persisted in naming difficult words despite the prior admonitions of the examiner. Almost naturally it should not be expected that a child of 6 years 7 months, his chronological age at the time of the examination, can give 28 words of the order of encyclopedia, evolution and luminosity in one minute. Yet these are some of the words he mentioned!

The above summary is nevertheless helpful in that it allows the reader to secure a better idea of the composition of K.'s mental age of 12 years 11 months. It is especially significant to note that he passes those tests which are viewed to be among the better items and which are less subject to coaching.

His Binet performance allows us to note further significant results. His vocabulary is equal to that of average 11 year old pupils. His auditory rote memory is good for his age. His comprehension of everyday situations involving language is superior for his age. Also, his visual representation and inductive reasoning are very superior for his age. Of greatest significance, perhaps, is the fact that on the basis of tests given it is found that approximately one child in a million obtains an IQ of 196, K.'s rating.

In order to feel somewhat safer in our complete judgment of K., it was thought desirable that he be given among the additional tests one or two of the non-language type. He was, therefore, given the Kohs Block Design Test, which was scored according to the revised system developed by Hutt.[2] His mental age was 12 years 3 months on the Kohs, as compared with the Binet M.A. of 12 years 11 months. This is a very favorable check. The following data characterizing the nature of his performance appear on the Kohs blank: very fine analysis; no trial and error placements; perceives relationships immediately and builds up pattern; careful auto-criticism; works deliberately.

He was also given the grade IV test of the Kuhlmann-Anderson. This level was chosen because it seemed to be best suited to K. He obtained a mental age of 10 years 6 months with an IQ of 158 (at this time of testing his C.A. was 6 years 8 months). This performance is a creditable one. Perhaps the

[2] M. L. Hutt, "The Kohs Block Design Tests. A Revision for Clinical Practice," *Journal of Applied Psychology*, June 1932, Vol. XVI, No. 3, pp. 298-307.

following notes on his Kuhlmann-Anderson blank partially account for the lower IQ of 158: attention and effort not uniform during the test; fatigued easily; when complete attention was secured, worked smoothly and easily without any apparent effort; total adjustment for the Kuhlmann-Anderson not altogether satisfactory. It should also be remembered that the IVth grade test requires reading. Though K. can read, as will be noted later, his relatively slow rate in reading because of the lack of adequate school experience therein is an additional possible factor in explaining the drop to an IQ of 158.

On the Goodenough Intelligence Test he secured a mental age of 9 years 9 months, also a favorable performance. Here again his failure to secure certain items and so a higher rating is due more to his relatively low C.A. and consequent incomplete development of the finer motor coordinations than to a lack of mentality. Also, many of the items on the Goodenough for which he does not receive credit are obtained mostly in performances indicating a special ability in drawing.

So much for our measures of general intelligence. Now, to what extent is K. making use of his mental capacities in his school subjects? Judged on the basis of the New York City median mental age limits for the various grades K.'s mental level is comparable to that of average pupils of the 7B grade. At present, however, he is only in the 2A grade, having entered the 1A at the age of approximately 6 years, and having skipped the 1B after having spent 6 months in the 1A.

In view of the level of his present mental capabilities, should it be expected of K. that he be able to do the work of the 7B grade? Obviously not, since he has not had the benefit of the necessary school training and experience. Should it be expected, then, that he do only the school work of his present grade? If so, what in terms of school achievements is representative of his very superior mentality?

Usually it is found that the superior child attains in achievement a level much above his grade placement but that he falls short of the achievements that would ordinarily be expected of an average youngster of a similar mental level. This is another way of stating that the A.Q. of the superior child is in most

cases below 100. So it is with K. He doesn't quite reach in his achievements the 7B grade but that he is certainly far above the work of the 2A, we shall soon see.

He was given several of the tests in the New Stanford Achievement battery. His comprehension of reading is excellent, being approximately equal to the median of pupils of the 6th year of elementary school. His comprehension is probably beyond this as can be seen by the fact that on the Stanford he secured 39 right out of 41 completions attempted. The point is that his rate of reading is somewhat slow because of the already mentioned insufficiency of drill and training.

His ability in arithmetical reasoning is equal to the median of pupils of the 6A grade.

It was expected that additional achievement tests would be administered to K., but at the time of the present writing records are not available.

Conversation with K. during the clinical examination revealed that he possesses an amazing informational background. This was not objectively evaluated. He shows familiarity with many items in geography, knows about various things in history, and is especially outstanding in his knowledge of astronomy. He indicated that he expects to become an astronomer.

Throughout the entire clinical examination K. showed no evidences of instability. His critical judgments were excellent. His use and comprehension of language were very good.

In attempting to propose an educational program for K. the psychologist finds himself confronted with a difficult proposition, and for many reasons. In the first place if the clinician is to be thoroughly scientific and candid, he must confess to a certain inability to secure an all-around psychological understanding of such individuals. They are too few in number to allow for many generalizations regarding their development and their direction. In the second place the marked hiatus between K.'s present level of intellectual maturity and his present levels of physical, physiological, social, and emotional development, produces a condition which is fraught with possibility for maladjustment. And in the third place there is the ineffectual nature, because of the rigidity of their curricula and their gen-

eral emphasis on the mass rather than the individual, of our publicly controlled schools that makes educational planning for K. a problem.

However, despite these difficulties inherent in the situation the following statements seem tenable. K. would benefit from placement in a school especially adapted to the education of superior children. One of the better progressive schools seems desirable. The choice of a progressive school as a suitable one for K. is made not because it is felt that "progressive" in itself assures Utopian adequacy in education. As a matter of fact, criticisms of theory and practice of our progressive schools easily present themselves.

The choice of a progressive school is made, however, because it is felt that here, better than in any other of our present day schools, can the curriculum be adjusted to suit K.'s particular needs. Here can be found an intellectual enrichment of experience so utterly lacking in most of our regular schools. Here can be found a method of pedagogy that invites pupil activity and spontaneity of expression. And here can be found an emphasis on the individual and his problems of growth and adjustment.

Suppose, however, that it is impossible for K. to enter one of the progressive schools and that consequently he finds it necessary to continue to attend the regular school. What then can be done for him? Of course, special attempts can be made to accelerate his progress through the grades or to enrich his curriculum or to combine factors of both of these plans into an educational program. These are theoretically desirable recommendations. In practice they present hardships. Some of these may be briefly indicated.

To place K. in a grade warranted by his mental age, though much beyond that warranted by his chronological age, would not be desirable. He would find himself in the midst of adolescent boys in the 7B grade much beyond himself physically, physiologically, emotionally, and socially. The undesirability of such placement is palpable.

To keep K. in the 2A, a grade level which he already surpasses not only in mental maturity but in school achievements

also, in order to provide him with companionship is equally inadvisable. It would not be conducive to the formation of the most favorable school attitudes or of the most desirable habits of application.

A plan of the following sort usually results. The very superior child is gradually accelerated through the grades. He is also given extra readings and assignments for purposes of enriching his background.

However, this plan dubiously and gratuitously assumes that enough teachers could be found who would care to be bothered with K.'s individual needs and interests. Then too, even if rapid acceleration were withheld, and it is hard to see how this could result, K. would, nevertheless, eventually find himself poorly placed for the previously indicated reasons by the time he reached the higher grades of the elementary school.

These are only some of the difficulties encountered in attempting to provide for K. in the regular school organization. Others often present themselves, although the immediate purposes of this article do not warrant an extended consideration of them.

It is hoped that the reader has gained insight into K.'s present mental functioning and into some of the complex circumstances which surround his growth and direction.

At the beginning of this paper it was noted that for considerations of the usual problems relating to precocious children there existed an extensive literature. Since K. showed no unusual problems in his social adjustment other than those generally found in very superior children, practically nothing has been said of such conditions as his association with playmates and his relationship to parents.

At a future date it may be possible to make a subsequent report on K., treating of his school progress, his interests, his vocational aspirations, and of his social adjustment.

A NOTE ON SILENT READING COMPREHENSION AND SUCCESS IN ACADEMIC ACHIEVEMENT IN A STATE TEACHERS COLLEGE

M. E. BROOM

State Teachers College, San Diego, California

ON A basis of certain experimental findings, Dickinson has asserted that reading plays an important rôle in determining scholastic success.[1] Book has written that "a student's success in every subject studied in college depends . . . very largely upon his ability to read and to get facts accurately and quickly from books."[2] It is the purpose of this discussion to report data bearing on the relationship between silent reading comprehension and academic achievement at the college level.

Two different groups of data are considered here. The first includes an aggregate of measures for 502 upper division students in the San Diego State Teachers College, assembled during 1928-1929; the second, measures relating to 171 upper division students in the college collected during 1929-1930. Academic success is measured in terms of average grade points earned per unit of credit completed in the college. The grade points per unit assigned to each grade are as follows:

A, excellent	3 grade points	D, poor	0 grade point
B, good	2 grade points	F, failure	—1 grade point
C, average	1 grade point		

It is possible to show the relationship of reading skill to scholastic achievement roughly, at least, by grouping the measures in accord-

[1] C. E. Dickinson, "A Study of the Relationship of Reading Ability to Scholastic Achievement." *School Review*, 33 (1925): 616-626.

[2] W. F. Book, "How Well College Students Can Read." *School and Society*, 26 (1927): 242-248.

ance with the placement of the student in the tenths for the total college population (separated by the deciles) on the reading division of the Thorndike Intelligence Examination for High School Graduates, Series of 1925-29.[3]

Table I summarizes the statistical information resulting from the application of this procedure. The data given in this table include the number of cases studied in each tenth grouping, the lower and the upper limits of the range of the average grade points earned per unit by all students in each such group, the mean, and the standard deviation for the frequency distribution of average grade points earned per unit by the individuals in each group. The significant fact shown in the table is that average achievement is greater in increasing degree in the groups having greater command of silent reading skills, if we accept average grade points earned per unit as a measure of academic achievement. It is worth noting that the ranges and the standard deviations vary much less in the upper half of the group and in the lower half of the group, considered as to separate divisions of the data, than do the means. Similarly, the standard deviations are quite uniform for the ten groups, which make the differences in mean achievement of relatively greater significance. The totality of impression resulting from a consideration of these data

TABLE I

Summary of Achievement Records of 502 College Students in Terms of Average Grade Points Per Unit for Ten Groups on the Reading Division of the Thorndike Test

Group	Number	Range		Mean	S. D.
		From	To		
10	89	0.24	3.00	1.89	0.61
9	52	0.43	2.93	1.73	0.55
8	50	0.51	2.61	1.65	0.45
7	55	—0.82	2.43	1.49	0.53
6	51	0.24	2.23	1.24	0.40
5	30	—0.52	2.30	1.25	0.65
4	42	—0.60	2.76	1.17	0.64
3	32	—0.97	1.92	1.06	0.57
2	36	—1.00	1.75	0.79	0.63
1	65	—1.00	1.92	0.76	0.66

[3]M. E. Broom and G. S. Bell, "The Thorndike Intelligence Examination." *High School Teacher*, 4 (1928): 188-189, 199.

is that reading comprehension is important for successful academic study, since the average student who demonstrates reading comprehension on a standardized test on his entrance into the college earns on the average a better academic record than does the student who is weak in reading comprehension.

The second group of data involves the use of three variables: (1) the total equivalent score yielded by the Thorndike test, measuring mental ability; (2) the reading score yielded by the reading division of the Thorndike test; and (3) the average grade points earned by the student per unit in all college work completed.

The total equivalent scores for the 171 subjects on the Thorndike test ranged from 22 to 100. The median was 59.19; the mean, 60.2; and the standard deviation, 14.86. The range of the Thorndike reading division scores included measures between 12 and 118. The median was 56.2, and the mean, 56.1. The standard deviation was 21.41. Average grade points earned per unit in this group ranged between 0.97 and 2.86, with a median of 1.36, and a mean of 1.30. The mean was described by a standard deviation of 0.60.

The relationship existing between the three series of measures for this group of 171 subjects may be made clear by the use of the correlation technique. The coefficient computed for the paired Thorndike reading division scores and average grade points earned per unit was positive 0.395; that computed for the paired Thorndike total equivalent scores and average grade points earned per unit, positive 0.513; and that for the paired Thorndike total equivalent and reading division scores, positive 0.775. The coefficient of partial correlation involving the relationship between Thorndike total equivalent and reading division scores with achievement in college study constant was positive 0.725. The coefficient involving the relationship between Thorndike total equivalent scores and college achievement with reading partialed out was 0.336. The partial coefficient expressing the relationship between Thorndike reading division scores and achievement with Thorndike total equivalent scores held constant was negative 0.006. The coefficient of multiple correlation expressing the relationship between Thorndike reading division scores against Thorn-

dike total equivalent scores and average grade points earned per unit was positive 0.775. The multiple coefficient describing the relationship between Thorndike total equivalent scores paired with Thorndike reading division scores and average grade points earned per unit was 0.807. The multiple coefficient involving the expression of relationship between average grade points earned per unit and the Thorndike total equivalent and reading division scores was positive 0.513. These several findings indicate that both mental ability as measured by the total Thorndike test and reading comprehension as measured by the reading division of the Thorndike test are important in successful college study. The variable or variables measured by the total Thorndike test seem to be somewhat more significant in aiding achievement in academic work in college than is reading comprehension.

It is possible to apply another technique to these data to evaluate quantitatively the contributions of the variables measured by the total Thorndike test and by the reading division of the Thorndike test to academic achievement to the college. Table II summarizes the data for this purpose. In this table the subjects were grouped in one of four quadrants in terms of their score placements with reference to the median values to the distributions of scores yielded by the total Thorndike test and by the reading division of the Thorndike test. The constants computed from the four resulting frequency divisions included for each quadrant grouping the lower and upper limits for the range of average grade points earned per unit by the subjects in the group, the mean, and the standard deviation. The measures in Table II verify the conclusions reached on the basis of the findings yielded by the correlation technique. Unfortunately the small groups in quadrants two and four make positive conclusions dangerous. It appears, however, that reading comprehension is valuable in aiding successful academic achievements in college although it is not as valuable for this purpose as whatever is measured by the total Thorndike test. This advantage will become clearer by a study of the means in Table II. When the quadrants are compared, the ranges are seen to be similar, except the lower limit of the range in quadrant three, and the standard deviations are similar in all the quadrants. The

significant facts are indicated by the differences between the means of each two adjacent quadrants. While the major average differences in academic success are a result of differences in mental ability, such differences result to a smaller degree from differences in silent reading comprehension.

The results of the two quantitative evaluations of the contribution of silent reading to college academic effects emphasize the importance of reading as a study tool. Apparently there is an obligation on the part of the college guidance officers in the matter of training weak readers, since improvement in reading skills results in improved success in college academic study.

TABLE II

A Quantitative Evaluation of the Contribution of Two Specified Factors to College Academic Achievement, 171 Cases

Thorndike Division Reading Score		Below the median (Md 59.19)		Above the median (Md 59.19)	
Above the median (Md 56.22)	Number=	15	Number=	71	
	Range=	0.16—2.18	Range=	0.38—2.71	
	Mean=	1.13	Mean=	1.75	
	S. D.=	0.59	S. D.=	0.51	
Below the median	Number=	71	Number=	14	
	Range=	—0.97—2.24	Range=	0.61—2.18	
	Mean=	1.06	Mean=	1.50	
	S. D.=	0.60	S. D.=	0.39	

A PSYCHONEUROTIC INVENTORY FOR CHILDREN BETWEEN NINE AND FOURTEEN YEARS OF AGE

FRED BROWN

Alfred Willson Children's Center, Columbus, Ohio

HISTORICAL INTRODUCTION

OBJECTIVE approaches to the study of neuroticism in childhood may be divided into three categories: (1) association test techniques, (2) analyses of intelligence test responses and (3) questionnaires. The first of these methods, devised by Jung (7) in 1910 and adapted for children by Lowell and Woodrow (17) has been neglected because of its unwieldy form, unadaptability to group study, and the difficulty involved in a statistical analysis of results.

The second method was based upon the assumption that neurotics would scatter over a wider range of year levels in the Stanford-Binet test than would normal children. Mateer's work (10) seemed to substantiate this view but the later investigations of Rosen (12), Wallin (16), Keys and Whiteside (8) and others seem to contradict these earlier findings and indicate that the psychometric approach is an unreliable indicator of neurotic tendencies.

The questionnaire method initiated by Hall in 1896 and used extensively in studies of child behavior and its anomalies seems to have proved more efficacious in the study of neuroticism. Woodworth's original list of 116 diagnostic questions (3) compiled in order to weed out potential neurotics among drafted men during the World War, was further developed and modified for use with children in their teens by Johnson (6), Cady (9) and Mathews (11). Studies by Slawson (15), House (5), and Terman (14), among others, corroborate the superiority of this approach over those previously attempted. Reliabilities of these various instruments range from +.369 to +.667.

Experimental results appear to indicate the need for a more reliable and valid psychoneurotic inventory which will be applicable to younger children. The present study represents an attempt to compile an instrument which will avoid the defects inherent in contemporary techniques.

NATURE OF THE GROUP

The group used in this study consisted of 1663 unselected children ranging in age from 9 to 14 years and including all grades from the fourth to the ninth inclusive. An effort was made to obtain a heterogeneous group with regard to habitat; therefore children from three Ohio cities, Cleveland, Columbus, and Dayton, were examined. Schools and grades selected for testing were as follows:

Cleveland	Columbus	Dayton
Fowler: 9	Everett: 7, 8, 9	Col. White: 7, 8, 9
Moses Cleveland: 4, 5, 6, 7, 8	West: 8, 9	Jefferson: 4, 5, 6
Roosevelt: 7, 8, 9	Medary: 4, 5, 6	
Taylor: 4, 5, 6	McGuffey: 4, 5, 6	

Chronological age ranged from 7 years, 7 months to 17 years, 6 months, but only 84 per cent of all cases are found below 9 years. A greater number was found to exceed the 14 year level, although no grade beyond the ninth was included. Our tables show that 15.49 per cent of our cases extend beyond the originally established upper limit; but only 2.64 per cent of the entire group, or 44 cases in all, exceed 15 years and 6 months. This excess of one and one-half years in chronological age is what we might expect in a group of this size and variability when we consider the different locations of the schools and the varying cultural status of the homes.

Intelligence level in terms of mental age and IQ was measured by means of the Morgan Mental Test. By comparing the IQ's with their calculated percentile values we noted that 30.94 per cent of cases fell below the 40th percentile rank. Morgan states that any IQ above 75 is good, which left us with 3.60 per cent

of children who were below average intelligence as measured by this test.

In order to determine the relationship between neuroticism and socio-economic status it was necessary to obtain as good a sample as possible of different and widely varying social and economic groups.

Our groups as determined by means of the Sims Socio-Economic Score Card (13) showed fairly discrete social and economic differentiation. Distributions by grade and school for the variables previously mentioned, including sex, were found to be well proportioned for the entire group.

The wide sampling obtained for these variables would seem to indicate that results found in this study should prove highly reliable when applied to children between nine and fourteen years of age, in the fourth to the ninth grade inclusive, and of either sex.

PROCEDURE

In selecting items for a more reliable inventory an effort was made to avoid faults committed by previous investigators. A careful analysis of literature pertaining to the neurotic child was made by sifting through the fields of psychology, psychiatry, pediatrics, sociology, psycho-analysis, and general medicine. Related studies of home environment, school behavior, social adjustment, and physical health of nervous children were examined and discrete symptoms tabulated upon filing cards. Final selection of symptoms was based upon the relative frequency of items in the clinical literature and the background and experience of the investigators. Practical studies were given preference over theoretical approaches. The first form of the inventory consisted of 176 questions distributed as follows:

School	11	Resistance to authority	3
Home	121	Sadistic tendencies	4
Physical	30	Hallucinations	2
Anxieties	5	Obsessive ideas	6
Sensitiveness	4	Superiority feelings	5
Social cooperation	11	Personal idiosyncrasies	9
Inferiority feelings	9	Indecision	4
Sleep disturbances	10	Temperamental	12
Nervous habits	10	Self-estimate	7
Escape mechanisms	7	Miscellaneous	6
		Total	176

This form was administered to 1663 school children between the ages of 9 and 14 years inclusive, from the fourth to the ninth grade of the public school.

The children were given a short introductory talk concerning the nature of the test and the necessity for serious coöperation. The fact that the test was not compulsory and that only those who wished to coöperate were wanted was strongly emphasized. The inventories were distributed after establishment of rapport, directions were read, and the children told to fill out the sheet as instructed.

The Morgan Mental Test and the Sims Score Card for Socio-Economic Status were administered after completion of the inventory. The former required only forty minutes and the latter approximately fifteen minutes for completion. At no time did the entire testing program require more than two hours.

Reactions to the inventory and accessory tests were very satisfactory. No signs of fatigue were observed or expressed in any of the groups examined. Many children expressed a desire to take more tests the next day, asserting that they enjoyed answering the questions.

Administration and supervision of the tests were controlled by two graduate students in clinical psychology and a junior in the College of Social Administration at the Ohio State University.

STATISTICAL TREATMENT OF THE DATA

Validity

The items were scored in accordance with their diagnostic significance in the clinical literature; i.e., if nail biting had been regarded as a neurotic item by authorities in this field, then an affirmative response was scored as a neurotic symptom. The practice adhered to by other investigators of regarding ambiguous responses (wherein both affirmative and negative responses were encircled) as indicative of atypicality was followed in our study. Approximately 75 per cent of all scoring was performed by the writer himself. The remainder of this arduous work was accomplished by an intelligent and efficient student in the department of social administration at the Ohio State University.

After all inventories had been scored, they were arranged in de-

scending order; i.e., those sheets with the highest total scores were placed above those with the next highest scores until those with the lowest scores were reached. The separate items were now ready for validation.

Two methods were resorted to in order to properly validate the items and select those which would be most differentiating. These methods, known as the method of Upper versus Lower Third or the U.L. method and the method of Overlapping or the O.L. method, are described by Lentz, Hirshstein, and Finch (9) who made a study of their relative reliabilities as compared with other methods of validation.

The U.L. method was used as follows: the highest third of the group (those with the highest total scores) and the lowest third (those with the lowest scores) were selected as the criterion group. The number of neurotic reactions on each item was counted separately for each group. Evaluation of each item was based upon the difference between the number of neurotic responses in the upper group and the corresponding number in the lower group. Items were selected by the excess of neurotically scored items in the upper group over those in the lower group.

In the Overlapping or O.L. method the total neurotic score for the individual was again considered the criterion score. The mean criterion score for all persons marking the item normally was calculated. The percentage of persons in the neurotic group whose criterion scores equalled or exceeded the mean for the normal group was computed and those items were selected which showed the lowest percentage of overlapping.

Table I represents the forty-six common items selected from the seventy-five best items in the U.L. and seventy-five best in the O.L. selections, together with a classification of the items. These items in themselves seem to present a sharply focused picture of the neurotic syndrome. In Table II, thirty-four additional items are presented which are not in statistical accord with each other by the two methods of validation but which are validated by one or the other of the two methods. The discrepancy between selections are not sufficient to invalidate the items.

TABLE I

Forty-six Common Items Selected from Seventy-five Best in U.L. and Seventy-five Best in O.L. Selections

Item Number	U.L.	Per Cent O.L.	Classification
75	300	15	Irritability in general
82	260	15	Anger reactions
65	256	18	Dissatisfaction with self
78	255	18	Sensitiveness
109	248	15	Transitory body pains
173	248	18	Insomnia
42	246	19	Awareness of conflict in personality
174	244	16	Sleep disturbance
74	239	18	Unstable temper
27	235	20	Being misunderstood
99	235	14	Fatigue
151	235	17	Recitation difficulty
63	230	14	Anxiety state
142	227	17	Financial worry
87	226	18	Sluggish blood circulation
81	225	20	Reaction to being teased
116	223	17	Reaction to biophysical stimuli
60	221	19	Indecision
61	215	20	Restlessness
115	215	18	Respiratory difficulty
157	215	13	Motor control
104	210	12	Fatigability
113	209	19	Reaction to biophysical stimuli
155	208	9	Accuracy in school work
47	207	18	Reaction to unreasonable demands
149	206	20	Reaction to school grades
95	204	15	Head pains
46	200	17	Memory
146	200	17	Parental domination
119	199	14	Motor control
162	197	14	Dreams
165	196	20	Excitement
105	194	14	Food idiosyncrasies
126	194	14	Escape from home environment
145	190	19	Negativism
127	190	14	Parental instability
100	189	18	Anaesthesias
83	186	15	Nervousness
152	186	15	Inability to concentrate
89	182	15	Vertigo
12	171	19	Rationalization for failures
101	170	16	Vomiting after meals
169	170	15	Attitude toward own dreams
97	168	16	Respiratory difficulty
129	168	18	Favoritism in home
106	167	18	Paroxysms of coughing or sneezing

It will be observed that the median excess of the upper versus lower scores for these thirty-four items is 130, with a range of

TABLE II

Classification of Thirty-four Additional Items Included in Revised Inventory and Selected by Both Methods

Item Number	U.L.	Per Cent O.L.	Classification
34	223	21	Introspection and social detachment
36	213	21	Lack of self-confidence
41	215	21	Persistence of anger state
48	74	19	Speech disturbances
53	107	17	Avoidance of competitive activities
54	129	18	Impulse to steal
58	202	21	Mental perseveration
73	132	8	Reaction to rebuke or reprimand
85	153	19	Frequent headaches
86	130	17	Cold extremities
94	114	16	Headache as an escape mechanism
96	50	13	Enuresis
103	189	18	Dejection
108	158	17	Twitching and jerking of limbs
110	114	19	Frequent micturation
117	130	17	Toothaches
118	104	8	Chronic fatigue
122	164	12	Gastric pains
123	122	17	Hysterical disturbance of sight or hearing
131	106	17	Social cooperation
132	96	18	Maternal strictness
133	137	10	Emotional instability in father
134	132	10	Suppression in home
138	111	10	Parental unfairness
139	154	15	Death dream involving father
140	147	19	Death dream involving mother
145	106	12	Vocal manifestation of parental dominance
147	51	9	Running away from home
148	85	11	Feelings of illegitimacy
156	118	14	Difficult school work
159	127	11	Cardiac disturbances
161	163	13	Food idiosyncrasy
167	237	21	Anxiety dreams
176	152	17	Anxiety dreams

from 51 to 223. The median percentage of overlapping is 17 per cent with percentages of overlap ranging from 8 to 21 per cent. It should be noted that fully 57.5 per cent of items constituting the final inventory were selected by two dissimilar methods.

Reliability

Since there was no objective criterion with which to test the reliability of our inventory, it was decided to use the method of

"internal consistency" or the "split-half" method for finding the coefficient of correlation between two halves of a test which has been applied only once. (4) This procedure yields the coefficient of reliability for the half test.

Three groups, composed of 100 each, were selected at random from the entire group in order to compute the reliability of the inventory. The correlation between odd versus even items was calculated for the eighty items selected by the U.L. and O.L. methods, the r for each group being calculated separately. In doing this, the Toops raw score formula for calculating the coefficient of correlation was used.

The reliability coefficients for the three discrete groups were as follows:

Grade IV ..... $r+.753$
Grade VI ..... $r+.866$
Grade VIII ..... $r+.803$

In order to compute the reliabilities of the whole test for each group the Spearman-Brown prophecy formula was utilized: (4)

This correction yielded the following reliabilities:

Grade IV ..... $r+.860\pm.010$
Grade VI ..... $r+.928\pm.006$
Grade VIII ..... $r+.890\pm.008$

The probable errors of our corrected r's were obtained by the use of Edgerton's tables.(2) The separate coefficients were averaged by the use of Fisher's Z formula and yielded a reliability coefficient of $+.896 \pm.007$ for the entire inventory.

Coefficients of correlation were computed for the relationship between neurotic score and chronological age, mental age, IQ, grade level, and socio-economic status. In calculating these relationships, we felt that no increase in reliability would be obtained by using a large group and that a random selection composed of every fourth case in the entire group would present a distribution similar to that manifested by the whole of our cases. We found the distribution of our sample group to follow that of the total group very closely. Our findings will be presented in the conclusions:

DISCUSSION OF DATA

An inventory of this type should impart qualitative or descrip-

tive data in addition to a quantitative score. Analysis of the various responses elicited interesting information with regard to the manifestations and concomitants of neuroticism in children. For example, we may notice certain differences between the two groups with regard to the home situation. The question, "Are you glad to get away from home once in a while?" (U.L. 80; O.L. 38 per cent) is insignificant because all children, as a rule prefer the uninhibited comradeship of members of their own groups to the petty demands of home and parents. However, when the question is asked, "Do you ever feel like running away from home?" a significant difference between neurotics and normals is observed (U.L. 194; O.L. 14 per cent). The latter exhibit a mild and passive rebellion against the demands of parents and older siblings; but there is always present an unverbalized recognition of the essential justice and inescapability of a situation which the average child accepts as an integral part of the culture pattern. For the neurotic child, it seems, the home environment assumes a different aspect. It is something unbearable, something to escape from definitely and permanently. Such escape tendencies seek to sever those bonds which sentiment has regarded as immutable and instinctive. When we ask, "Have you ever run away from home?" (U.L. 85; O.L. 9 per cent) the evidence appears to confirm our theory.

Psychoanalysis has stressed the importance of dreams as complex indicators. Our data indicate some relation between dreams and neurosis. Children we find do as a rule relate their dreams to their parents. There is, however, a distinct difference between the two groups in the attitude taken toward the dream content, which might indicate that such dreams possess an unconventional manifest content, possibly of a sexual nature.

In answering the question "Do you ever feel ashamed of your dreams?" the neurotic group shows an excess of 170 by the U.L. and an overlapping of only 15 per cent by the O.L. method. In connection with this the O.L. percentages for questions "Have you ever dreamed that your father died?" and "Have you ever dreamed that your mother died?" are 15 per cent and 19 per cent respectively. The anxiety dream, exemplified in the questions "Have you ever dreamed that you were locked in a room

and could not get out?" and "Did you ever dream that you are running away from someone who is trying to harm you?" shows a small percentage of overlapping. It may justifiably be concluded that the dreams of psychoneurotic children offer a clue to the syndrome as a whole and would repay further research.

A predominance of physical symptoms is found, which confirms medical investigations in this field. Headaches, enuresis, migrainous headaches, vertiginous sensations, respiratory difficulty, chronic fatigue, temporary anaesthesias, vomiting after meals, sickness at the sight of food, fits of coughing and sneezing, transitory pains and frequent micturation predominate in the neurotic child.

Significant differences with regard to the home situation, school adjustment, social adaptation and self-estimate were discovered; whereas certain symptoms often mentioned in the literature, such as left-handedness, blushing, being the oldest or youngest child in the family, sleeping in the same room with parents, and ability to tolerate pain were found to be statistically invalid.

CONCLUSIONS

1. The inventory compiled for this study possesses a high reliability, which should render it useful in research problems and for clinical purposes. A reliability of $+.896 \pm .007$ was obtained for the entire inventory, which was found to be considerably higher than that for any other inventory in use at the present time.

2. No correlation between chronological age and neurotic score was revealed, indicating that the instrument is independent of age level. ($+.009 \pm .03$)

3. The difference between the mean neurotic scores of boys and girls was found to be statistically insignificant. The inventory appears to be equally applicable to both sexes.

4. Neurotic score is uninfluenced by intelligence level. This conclusion is borne out by a correlation of $-.111 \pm .03$ with M.A. and $+.013 \pm .03$ with IQ.

5. Neuroticism does not predominate in any particular social or cultural group. Situations engendering nervousness in children may be encountered in any stratum of society. A correlation

of —.118 ±.003 between socio-economic status and neurotic score was obtained.

6. A slight tendency for children in the fourth grade to make higher neurotic scores than those in higher grades was observed. The correlation between grade level and neuroticism was found to be —.370 ±.03. Further evidence is required in order to determine whether increased demands made upon the child at this stage of the educational process is responsible for the appearance of neurotic behavior.

7. No significant differences between age-grade standing and neuroticism were found. Reliable differences in the intelligence of neurotic and normal children in the fourth grade were obtained: The critical ratios $\frac{(\text{Diff})}{(\sigma\text{diff})}$ for M.A. and IQ between the two groups were 2.73 and 2.17 respectively. This conclusion is based upon a paired study of 20 neurotic and 20 normal fourth grade children. The C.R. for grades VI and VIII were .13 and .37 respectively for IQ (N=24 normals and 24 neurotics).

8. Children who make high neurotic scores show reliable differences when compared with normal children with regard to dreams, home situation, physical symptoms, school adjustment, and general social adaptation.

BIBLIOGRAPHY

1. Cady, Vernon, "The Estimation of Juvenile Incorrigibility," *Journal of Delinquency Monograph No. 2* (April, 1923).

2. Edgerton, Harold, "A Table for Finding the Probable Error of R Obtained by the Use of the Spearman-Brown Formula (N-2), *Journal of Applied Psychology, 14,* 296-302.

3. Franz, S. I., *Handbook of Mental Examination Methods,* New York, 1919, p. 193, The Macmillan Company.

4. Garrett, H. E., *Statistics in Psychology and Education,* Longmans, Green and Company, 1931, pp. 268-269, 271.

5. House, S. D., "The Problem of Childhood," *Journal of Delinquency, 12,* 25-42 (1928).

6. Johnson, Buford, "Emotional Instability in Children," *Ungraded, 5,* 73-79 (1920).

7. Jung, C. G., "The Association Method," *American Journal of Psychology, 21,* 219-269 (1910).

8. Keys, N. and Whiteside, G. H., "The Relation of Nervous Emotional Stability to Educational Achievement," *Journal of Educational Psychology, 21,* 429-44 (1930).

9. Lentz, T. F. Jr., Hirshstein, Bertha, and Finch, J. H., "Evaluation of Methods of Evaluating Test Items," *Journal of Educational Psychology, 23,* 344-50 (May, 1932).

10. Mateer, Florence, *The Unstable Child,* D. Appleton and Company, (1924).

11. Mathews, Ellen, "A Study of Emotional Stability in Children by Means of a Questionnaire," *Journal of Delinquency, 8,* 1-40 (January, 1923).

12. Rosen, E. K., "A Comparison of the Interests and Educational Status of Neurotic and Normal Children in Public Schools," Teachers College, Columbia University, *Contributions to Education,* No. 188, (1925).

13. Sims, V. M., *The Measurement of Socio-Economic Status,* Public Schools Publishing Company, 1928.

14. Terman, L. M., *Genetic Studies of Genius,* Vol. 1, Stanford University Press, p. 648 (1925).

15. Slawson, John, *The Delinquent Boy—A Socio-Psychological Study* (Richard G. Badger), The Gorham Press, Boston, 1926.

16. Wallin, J. E. W., "Intelligence Irregularity as Measured by Scattering on the Binet Scale," *Journal of Educational Psychology, 13,* 140-152 (March, 1922).

17. Woodrow, H. and Lowell, F., "Children's Association Frequency Tables," *Psychol. Monographs, 22,* p. 97 (1916).

MODIFICATION OF THE KOHS BLOCK DESIGN TEST

GRACE H. KENT

Danvers (Mass.) State Hospital

1. INTRODUCTION

THE Kohs Block Design Test was first offered in 1920 (1) and was presented more fully in 1923 (2). By making use of a popular toy which consists of painted cubes having different colors or color combinations for the six sides of each cube, Kohs gave the clinic a performance test which is unusually attractive to the juvenile subject. There are seventeen designs to be built up with the cubes, the number of blocks required for a given design ranging from four to sixteen. The actual dimension of the cube is slightly over one inch; but the Kohs patterns, which are painted on white cards, represent the cubes on a half-inch scale, thus requiring the subject to enlarge his image of the pattern as he sees it.

In the color cube test devised by Maxfield (3) and later standardized by Hutt (4) these same painted cubes are used for assembling five four-block designs, the pattern being presented as a built-up square of four cubes instead of being painted on a flat surface. In the last two tasks the pattern is withdrawn after a ten-second exposure, so that the subject is required to reproduce it from memory.

The Kohs test was first used by the writer in 1922 at Worcester State Hospital, and was found to possess the following strong features:

1. It is an outstandingly attractive performance test, employing a toy which appeals instantly to children and which is scarcely less fascinating to many adults. It is a test which evokes compelling interest on the part of the subject, in the large majority of cases, and which therefore commands spontaneous cooperation.

2. The test can be presented by pantomime, wholly without

the use of language. It can therefore be used for deaf children, and to some extent for immigrants.

3. The cost of the materials is trifling. The set of sixteen cubes can be obtained at department stores for about forty cents, and the designs can be painted at home.

4. The complete outfit is so light and compact that it can be added to the equipment of the traveling clinic even when the more cumbersome performance tests have to be omitted.

5. The test covers a wide range of discriminative capacity, from seven years to the superior adult level.

In spite of these features, however, the test as standardized by Kohs has not been widely adopted for routine use in clinics. The factors which have prevented its adoption in at least six institutions are as follows:

1. The performance of the seventeen tasks occupies a full hour, almost as much time as can usually be allowed for the entire examination. The test thus requires far more time than can reasonably be spared for a single test of limited scope.

2. The scale is stronger at the upper than at the lower end, for which reason it is better suited to industrial than to clinical needs. The great need in the clinic is for lower-level tests, as opposed to tests which show fine discrimination for varying degrees of ability at the superior level. The difficulty of the Kohs tasks is increased, and the clinical serviceability of the test accordingly diminished, by the fact that the patterns are made on a reduced scale as compared with the blocks which are used for copying them. This difference in size between the cubes and the designs is especially confusing to subjects of low mentality and to intelligent adults who have had no experience in solving puzzles. There are subjects of all ages who entirely fail to grasp the similarity in form and color. Although the discrepancy in size is not seriously disturbing to a subject who rates high in the test, for one of low achievement it may make all the difference between success and failure, and thus the test discriminates against the very subject for whom it is most needed.

3. The scoring system used by Kohs depends in part upon a count of the subject's moves, a plan of evaluation which is open to serious objection. In the first place, the clinical subject is fre-

quently too sensitive to tolerate being closely watched by the examiner. Many subjects react so strongly to close observation as to invalidate the test as a measure of their actual capacity, and at best such observation is not conducive to the poise which the subject requires for efficient performance of a task having a time limit. And in the second place, it is usually difficult if not impossible to make a passably accurate count of the subject's moves. This second objection to move counting will receive further notice in a later section.

In order to overcome these defects so as to make this very attractive test serviceable in the clinics of Worcester State Hospital, the writer modified it in 1923. The test was shortened by reducing the number of tasks from seventeen to eight, and the selected designs were enlarged to the scale of the cubes. Later four very simple designs were added to the series, in order to strengthen the scale at the lower end. It is appreciably weakened at the upper end both by the enlargement of the designs and by the omission of the most difficult task of the Kohs series; however, the scale is still discriminative at the adult level.

The test thus modified was intended for personal use and not for publication. After it had been given out privately to several clinics it received mention in the Bronner-Healy-Lowe-Shimberg Manual of Individual Tests and Testing, pages 147 and 210 (5). Since the appearance of this book in 1927 the writer has received many inquiries concerning the test and has made a practice of giving the revised norms to all inquirers. It is in response to requests from persons who are using the test in this form that it is now offered to the psychological public as a supplement to the color cube tests already published. It is probable that the Kohs test will still be preferred at the superior level whenever the time can be spared for a one-hour examination. The Maxfield-Hutt test, being much stronger in respect to standardization, should be preferred at the age levels which it covers. In the writer's experience, however, this test is not discriminative below eight years nor above twelve years. In examining a subject whose ability as measured by the test is wholly unknown, it is convenient to use a test having a wider range of discriminative capacity.

Inevitably the long continued use of a test brings to light certain defects which might have been corrected had they been recognized in the early stages of development. The possibilities of this color cube test have by no means been exhausted. Unfortunately the Maxfield-Hutt test came to the writer's notice after the edition here described had been in use over two years and after its standardization was already too far advanced to permit further modification of method. By taking full advantage of Maxfield's plan of presenting the pattern it would be possible to make the test discriminative at five years, and possibly at four years. The first task might well be a row of three or four solid red blocks, copied from a pattern built up in full view of the subject; and the second task might be a four-block square of solid yellow. At the upper end of the scale the difficulty might be increased either by using a larger number of cubes for a single task or by following the Kohs plan of using reduced-size patterns.

The advantage of having a clinical test discriminative from early childhood to the superior adult level will hardly be questioned, because in many cases the subject's social history is not on record and his ability is wholly a matter of conjecture. But although it is highly desirable that the tasks of the test should form a long series of graduated difficulty, the norms should be offered in the form of several overlapping scales rather than one continuous scale. The upper and lower ends of the series are mutually exclusive in their suitability, and therefore it should never be necessary to use the series in its entirety for any given subject. A color cube series might well be divided into three separate scales, independent as to norms while yet having some tasks in common.

The test here offered is only an approach to this ideal. The twelve tasks are discriminative from six years to adulthood, and the norms are given in the form of two independent scales.

2. MATERIALS FOR THE TEST

1st. The Embossing Company's Color Cubes, sixteen blocks. A set of these cubes can be purchased at local toy shops, or can be ordered from R. H. Macy and Company, New York City.

2nd. A set of twelve painted designs, illustrated in Figure 1. The designs numbered 1 to 8, most of which are taken directly from Kohs, constitute the Diagonal Series; and in contradistinction to these, the four designs designated as C, D, E, F, have been called the Straight Series.

These designs can be obtained from The Stoelting Company, Chicago. But inasmuch as most clinical workers prefer to make their own materials whenever possible, full instructions are given here for painting the designs.

They may be painted upon white bristol-board, or upon any heavy white drawing paper having a smooth but unglazed surface. This may be cut into cards about six inches by eight. It is well to have a set of the cubes at hand and to build up whatever design is to be drawn or painted, so as to work from a full-size copy.

Graph paper ruled in tenths of an inch may be used for finding the main points of the design. As the cubes are slightly over one inch in size, it is well to allow eleven-tenths for the square which represents a single cube. This is not exactly to scale, but is sufficiently so from the subject's point of view. For greater accuracy, millimeter cross-ruled paper may be used.

The graph paper is fastened to the card with paper clips, and the desired points are indicated by pricking through with a fine needle. The perforations should be barely large enough to be clearly visible. It is much easier to make a pinhole for each corner point than to omit the few which are not needed.

Guided by the pinholes, connect the significant points by ruled lines drawn lightly with a hard pencil, as illustrated in the drawing of design 5 which is marked with an asterisk. The pencil lines should show the full outline of each area of a given color, but not the outline of individual squares which are alike in color. (In this drawing the three unused perforations are indicated by circles. These extra pinholes are harmless, if made small enough to be fully covered by the paint.) The outside lines for designs 4, 6 and 8 should be heavy enough to show a clear demarcation between the white of the pattern and the white margin of the card.

Oil paints or water colors may be used, as preferred, and they

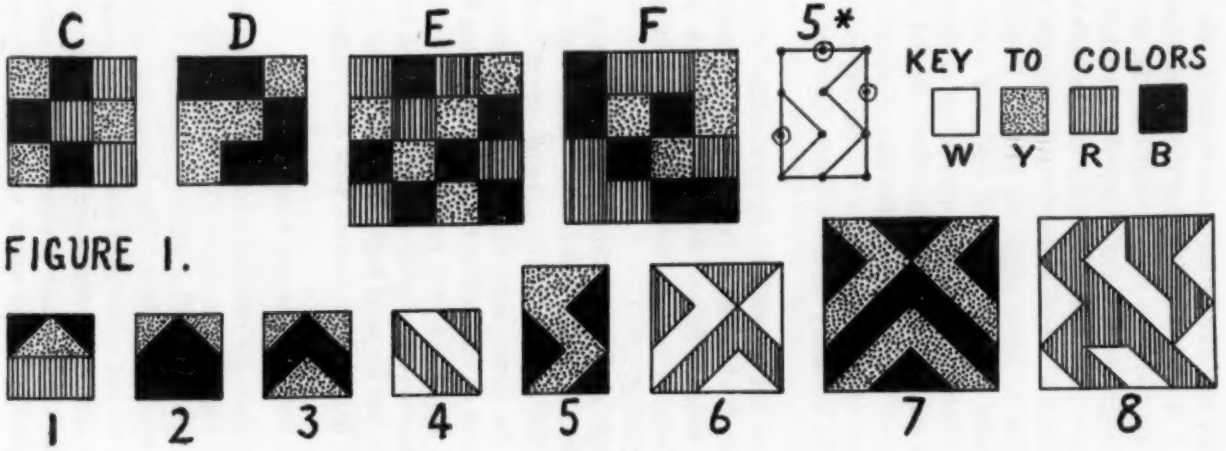

FIGURE 1

Illustration of the twelve designs. Blue is indicated by solid black, red by parallel lines, and yellow by irregular dots.

should be carefully matched to the colors of the cubes. (For research purposes it is recommended that the cubes themselves be repainted with quick-drying enamel and that the same paints be used for the designs, so as to insure a perfect match.) A small brush of soft bristles is used, slightly stiffer than camel's hair. An inexperienced person will find it worth while to begin by painting a few triangles on waste paper, for practice in painting the corners. The paint should be rather thick, as otherwise two coats may be necessary. The white areas of the designs are left unpainted. The yellow paint is applied first, the red next, and the blue last. In no instance should the darker paint be applied until the final coat of the lighter color has become passably dry. Usually it will be necessary to allow three days for the three colors.

A set of designs can be made more serviceable by giving the cards a light coat of white shellac, to prevent the margins from becoming easily soiled. Yellow shellac must not be used. In laying on the shellac it is important to brush evenly, with the grain of the cardboard.

The twelve cards are kept in serial order by making them up into a booklet, held together by notebook rings. Before punching the holes for the rings, it is well to reinforce the inner edges of the cards, on the wrong side, with a strip of gummed paper or cloth. The leaves of the booklet should be separated by sheets of waxed paper, because it may require weeks or months for the varnish to become thoroughly dry. If protected by waxed paper, the booklet can be used freely long before the leaves are dry enough to be left in direct contact with one another.

3. INSTRUCTIONS FOR PRESENTATION

As stated above, there are two scales having independent norms but having some tasks in common. The upper scale is based wholly upon the eight diagonal designs, presented in serial order until the subject reaches his limit. The lower scale includes all credits which the subject achieves in the straight series, plus any additional points which he can score in the diagonal series. Either scale may be used, or both scales, for subjects rating from eight to eleven years; but the upper scale is usually preferred

for those who rate upwards of ten years, and the lower scale for those who rate below ten.

The examiner can decide almost automatically which scale can be used more appropriately for a given subject. Designs C and D are presented first, followed by the diagonal series. The subject who completes the eight diagonal tasks is invariably rated by the upper scale, and it is therefore unnecessary to have him complete the straight series. But the tasks E and F should usually be given to the subject who fails to complete the diagonal series, even when they are not needed for his score, in order that the test may leave him with a sense of achievement rather than a sense of failure.

Throughout the examination try to keep the subject at his ease by methods adapted to the individual subject. A very immature person may respond well to extravagant and undeserved praise, while a more intelligent subject might feel insulted by the same treatment. A nervous and diffident subject can frequently be reassured by a casual and rather nonchalant attitude on the part of the examiner. It is recommended that the examiner have at hand a book which he can pretend to read, so that the over-sensitive subject may feel less keenly conscious of being watched. Occasionally the examiner may even pretend to copy

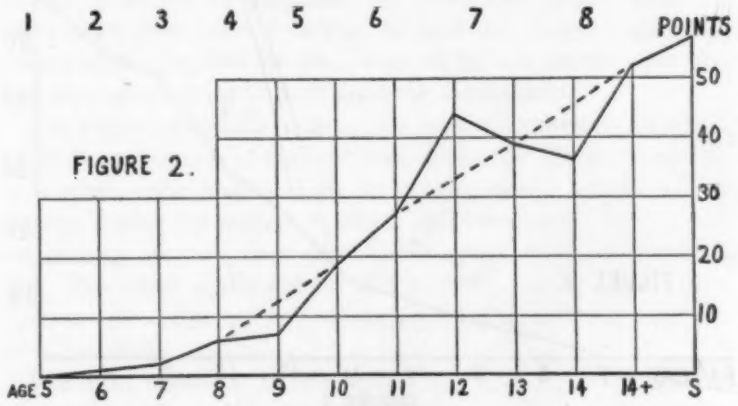

FIGURE 2.

FIGURE 2

Median Scores for Upper Scale. Broken line indicates artificial alterations for norms.

from the book, while actually taking notes on the subject's mode of procedure. Inasmuch as the subject is usually too busy to follow the examiner's eyes, it is possible to make all necessary observations without letting him see that his movements are being closely scrutinized.

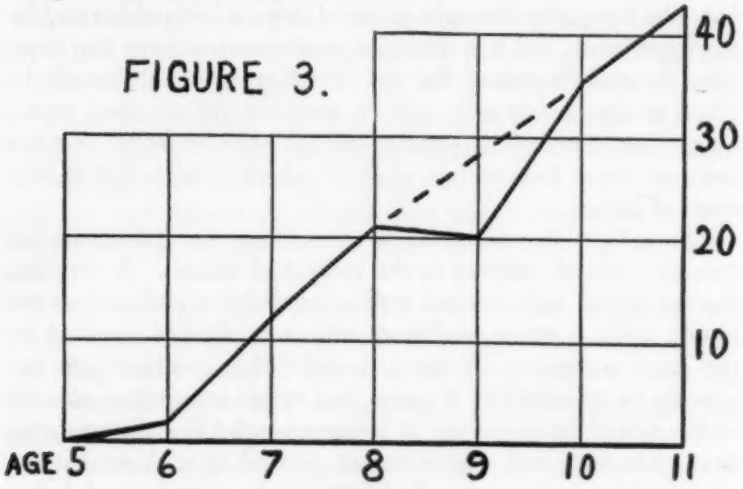

FIGURE 3

Median Scores for Lower Scale.

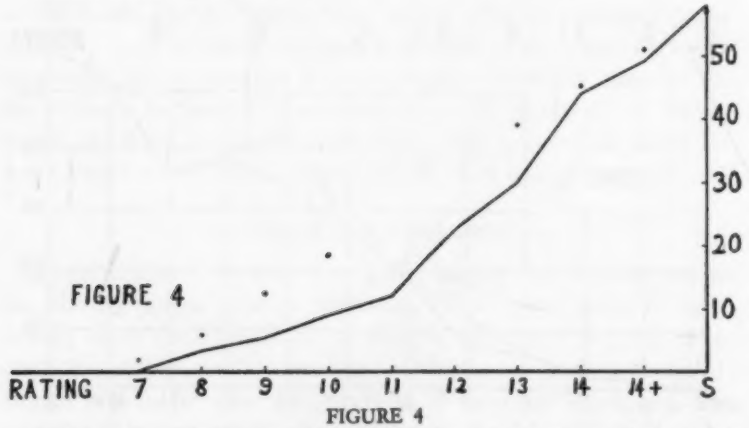

FIGURE 4

Median Scores of 250 State Hospital Patients, grouped according to their ratings by language tests. The dots indicate the age-curve derived from normal children, altered at ages 9, 12 and 14.

The stop watch should be held under the edge of the table, and should be operated as noiselessly as possible. Of course it would be unfair to the subject to withhold from him the fact that his performance is being timed, but it should not be forced unnecessarily on his notice. Most subjects work rapidly of their own accord and many of them are keenly aware of the stop watch. Only occasionally is it necessary to request a subject to see how quickly he can complete the task.

Examiner and subject should be seated on opposite sides of the table, preferably with the light coming from the end of the table. The lighting of the room requires special attention, whether natural or artificial, because there are certain angles from which the light may be so reflected as to make cubes and designs appear to be of one solid color.

When showing the blocks to the subject, turn them over and around so as to let him see the different sides and call his attention to the fact that the cubes are all alike. Give him nine blocks, keeping the others out of reach. Show him design C, and request him to place the cubes on the table so as to make them look like the pattern. Explain as fully as necessary that the cubes are to be arranged in the form of a square and that the top of this square should look exactly like the pattern. The subject must not be permitted to cover the design with the blocks. In some cases it is well to hold the design slightly inclined, explaining that the blocks are to be laid on the table and that the pattern must be kept where it can be seen.

The subject who fails to grasp the requirements may be given a full demonstration of the first task, design C. Point repeatedly to a given cube and then to its corresponding square in the pattern, giving the subject as much assistance as he requires for completing the task, then turn the design around for a second trial. No credit is allowed for this performance, but the subject receives full credit for any subsequent task which he can perform unaided.

The same procedure is followed, if necessary, at the beginning of the diagonal series. The subject may be given a demonstration of the first task, after which he may be allowed to try again for practice. He loses credit on this design, but may yet be able to score on the next task.

When presenting the test to a subject having no understanding of any language which the examiner can use, it is advisable to build up several simple designs for the subject to copy from the blocks. Design C should be reserved until the nature of the task is well understood.

For each task the subject is given the correct number of cubes. Between performances the blocks are shuffled carelessly and presented in no particular order, except that no two blocks should be placed so as to form a part of the design; and at least one block should be right side up.

The second task, design D, may be omitted when it is clearly apparent that the subject is ready for the diagonal series without further practice in handling the cubes.

As indicated in the scoring table, each task has its time limit beyond which no credit is given for a successful performance. The subject, however, should not be informed that there is a time limit. After passing the limit he should be allowed to complete the task unaided whenever there is a reasonable chance of his ultimate success, or he may be given just enough assistance to enable him to finish.

A subject who is too ready to give up should be encouraged to keep on trying, unless the examiner believes the task to be altogether beyond his ability. When the subject overlooks an error and is satisfied with an incorrect performance, well within the time limit, it is permissible to request him to look his work over carefully and see if he can find anything wrong with it. The time is recorded at the end of the entire performance, no deduction being made for the time lost by thus calling his attention to the error. If the subject can correct his work without further assistance he may receive full credit on the ground that the error is sufficiently penalized by the loss of time.

Two consecutive failures in the diagonal series,—inherent failures, indicating actual inability to complete the tasks—may be taken as proof that the subject has reached his limit, and the examination is brought to a close by returning to designs E and F of the straight series. But a failure to score merely by reason of exceeding the time limit does not offer any ground for considering the examination finished, because the practice effect of

a slow performance may enable the subject to accomplish the next task within the time limit.

Any subject who is to be rated by the test is technically entitled to try each task, so long as there is a possibility of his achieving an additional point. A zero score for the entire straight series should not be accepted as final unless the subject has been permitted to try at least one task of the diagonal series. It is unusual but by no means unprecedented for a subject who cannot copy the straight designs to score a few points on the smaller designs of the diagonal series.

No rating can be considered fair if assistance has been given within the time limit of any task, nor if the series as a whole has been brought to a premature close. However, the advisability of bringing the test to full completion must depend upon the reaction of the subject. It is better to report an examination as unfinished than to let the subject's self-confidence suffer a serious shock because of his inability to perform a seemingly simple task. The person tested is more important than the rating by any test,—something to be kept especially in mind when the test in question is one of very limited scope.

4. RECORDING AND EVALUATION

The time of each performance is measured by a stop watch. The scoring is based wholly upon speed on the ground that this is the only variable which can be measured accurately and objectively. The time record should always be supplemented by descriptive notes.

Tables I and II show the values which have been assigned for performance of any task within specified time intervals. Four scoring systems were tried out before this one was adopted, two of the discarded systems being more elaborate than this and the other being simpler and more arbitrary. The scoring scale for the Straight Series is quite artificial, but that for the Diagonal Series is based fundamentally upon statistical findings, altered barely enough to make it convenient for routine use.

Tentative norms are offered in Table III. It should be clearly understood that the norms are very crude and inadequate, as explained in the section on Standardization. Figure 2 shows to

what extent the curve of medians has been altered in order to make the norms usable.

TABLE I

Scoring Values for Diagonal Designs, According to Specified Time Intervals as Given in Minutes and Seconds

Time	Designs 1	2	3	4	5	6	7	8
						Points		
1″ to 15″	3	4	6	7	9			
16″ to 30″	2	3	5	6	8	11		
31″ to 1′	1	2	4	5	7	10	13	15
1′ 1″ to 1′30″	1	1	3	4	6	9	12	14
1′31″ to 2′	—	1	2	3	5	8	11	13
2′ 1″ to 3′	—	—	—	2	4	7	10	12
3′ 1″ to 4′	—	—	—	—	—	6	9	11
4′ 1″ to 5′	—	—	—	—	—	—	8	10
5′ 1″ to 6′	—	—	—	—	—	—	—	9

TABLE II

Scoring Values for Straight Designs, According to Specified Time Intervals as Given in Minutes and Seconds

Time	Designs C	D	E	F	
					Points
1″ to 30″	4	5	6	7	
31″ to 45″	3	4	5	6	
46″ to 1′	2	3	4	5	
1′ 1″ to 1′30″	1	2	3	4	
1′31″ to 2′	—	1	2	3	
2′ 1″ to 2′30″	—	—	1	2	
2′31″ to 3′	—	—	—	1	

TABLE III

Tentative Norms

Age	POINTS Upper Scale	Lower Scale
6	1	1–8
7	2–4	9–15
8	5–8	16–23
9	9–14	24–31
10	15–22	32–38
11	23–29	39 up
12	30–36	
13	37–42	
14	43–48	
14	49–54	
S	55 up	

5. STANDARDIZATION*

The total number of subjects included in the norms is 387. The first 132 records were obtained in the schools of Leicester, Massachusetts, in 1925. Records from 125 Danvers children were added in 1931, including 41 pre-school children who were examined for school entrance. All other records, 130 in number, have been picked up at random whenever the opportunity has offered itself, a large proportion of them having been contributed by college students and professional persons who have had occasion to visit the laboratory. The ages of these adult subjects have not been ascertained, but it is a safe guess that most of them were in the early twenties. The ages of juvenile subjects are reckoned from the nearest birthday.

Tables IV and V show the median scores at each age level for the upper and lower scales respectively, also the range of scores and the number of cases included. With a few accidental exceptions, records obtained from children under twelve years of age are used in both scales.

The number of cases is far too small to justify any considerable confidence in the norms, and it will be noted further that the scale is especially weak at the center as compared with the ends. The material includes 41 records from five-year children and 133 records from subjects over fourteen, thus leaving only 213 records for the ages from six to fourteen.

TABLE IV

Median Scores for Upper Scale, with Number of Cases and Range of Scores at Each Age Level. 383 Cases

Age	5	6	7	8	9	10	11	12	13	14	14+	S
Number of cases	41	19	23	23	20	24	29	20	25	26	47	86
Range of scores	0-13	0-21	0-36	0-56	1-45	4-56	1-56	3-58	7-57	7-60	6-62	29-65
Median score	0	1	2	6	7	18.5	27	44	39	36	52	57

*In collecting and assembling material for norms, assistance has been received from the following persons: David Shakow, Alice Paine, E. Frances Wells, Katharine Simpson, Johanna Hagg, Farrior Brown, Mary McJennett, Elizabeth Alexander, Dorothy McLeod, Alice Schoenfuss. This aid is gratefully acknowledged.

TABLE V

Median Scores for Lower Scale, with Number of Cases and Range of Scores at Each Age Level. 181 Cases, Under Twelve Years

Age	5	6	7	8	9	10	11
Number of cases	41	23	23	22	20	23	29
Range of scores	0–13	0–31	0–49	11–75	9–61	12–76	13–76
Median score	0	2	12	21	20	35	43

The median scores are presented also in graphic form, Figures 2 and 3. The artificial alterations in the age curve are indicated by broken lines.

The rise with increasing age, however irregular, is significant for so small a number of cases. Except at the ages 9, 12 and 14, the medians are accepted as they stand. The irregularities are smoothed out by simply omitting these three levels.

Both the Leicester records of 1925 and the Danvers records of 1931, studied independently, show low achievement at nine years as compared with eight, and high achievement at twelve years as compared with thirteen. The distortions of the curve at nine and twelve are apparently due to this coincidence.

The drop at fourteen years, however, cannot be considered wholly fortuitous. Here the discrepancy is only what may be expected when a test is standardized in the elementary schools. As we proceed upward in the age scale it becomes increasingly difficult to find a representative group of subjects. In all the tests standardized by the writer, it has been necessary to employ advanced high school students well above fourteen years of age in order to obtain a score appreciably higher than the thirteen year score. Many of these tests have yielded a lower score at fourteen than at thirteen, thus requiring artificial alterations in order to make the tests useful beyond the thirteen year level. Attention has been called to this in another paper (6).

6. STATE HOSPITAL DATA

As a side-light on the practical serviceability of the test, it seems worth while to present in brief some results obtained from psychotic patients.

In Danvers State Hospital it is customary to employ psychometric tests as auxiliary to the routine mental examination of newly admitted patients. For literate patients rating from eight to fourteen years, the language test most commonly used is the Kent-Shakow written battery of seven units, as standardized without a time limit (7). This battery yields seven independent "mental age" ratings, the median of which is accepted as the final rating by this part of the examination.

The median rating by this battery has been found useful for calibration of tests which are not adequately standardized. Subjects are grouped according to their median ratings, exactly as if these ratings represented their actual ages; and within each group of a given mental level, the scores achieved in the unstandardized test are arranged in order of rank. The median score, for each mental level, is used as a check on whatever crude norms are available.

The color cube test here described, although not used as a routine measure, has been given to nearly 800 patients during the past five years. At present writing (June, 1933), the files contain 237 color cube records obtained from patients who received ratings by the seven-unit battery, their median ratings ranging from eight years to fourteen-plus. This range is extended at the ends by including a few who received a rating of seven by Stanford-Binet and a few who received high ratings by Army Alpha, making 250 cases in all. The color cube scores, for the upper scale only, have been calibrated by these ratings.

Table VI shows the median color cube score for each mental level thus established, with the number of cases and range of variation at each level. Figure 4 shows these medians in graphic form, for comparison with the tentative norms. The dots represent the altered age-curve derived from school children, not in all cases the actual medians.

It will be noted that the achievement of these patients in the color cube test is much lower than their achievement in the language tests, except at the upper levels. Two possible explanations are offered for this: (1) that adults are weak, as compared with children, in a performance test which appeals to the toy-interest; and (2) that psychotic subjects achieve higher scores in

tasks having no time limit than in tasks scored by speed.

The two groups of records, of course, are not comparable. No records are available from non-psychotic adults of ages corresponding to those of the patients, nor are there any battery records to be used for calibrating the color cube records obtained from normal children. Without such data, it is impossible to offer any conjecture concerning the relative importance of the two factors cited.

TABLE VI

Color Cube Scores, Upper Scale, Calibrated by Language Tests. State Hospital Patients, of Ages from 14 to 59. 237 Cases Rated by Kent-Shakow Written Battery, with 13 Additional Cases Rated Seven by Stanford-Binet or Superior by Army Alpha

Rating by language test	7	8	9	10	11	12	13	14	14+	S
Number of cases	5	4	16	36	25	48	59	41	8	8
Range of scores	0–13	1–8	0–24	0–48	0–55	2–58	0–57	9–62	28–55	37–63
Median score	0	3.5	5.5	9	12	22.5	30	44	49	58

Inasmuch as further use is to be made of the state hospital material, it is not necessary at this time to analyse it more closely. Reference to it is included here by way of proof that color cube ratings are even less trustworthy for adults than for children. In this connection it should be noted that the adult scores, at each mental level which is well represented numerically, range all the way from very low to very high. (Table VI.) It is quite evident that for psychotic patients and for adult immigrants the test should be used with the utmost caution, for qualitative rather than quantitative interpretation.

7. SERVICEABILITY AND LIMITATIONS

This edition of the color cube test has seen ten years' service in a few clinics. At present writing only one performance test is more frequently used at Danvers State Hospital—the Kent-Shakow form board series, another test which is discriminative from six years to the superior adult level (8). The color cube test does not rank with this form board series for general

serviceability in the clinic. The form board task is self-corrective, as the color cube task is not. A self-corrective task is more easily comprehended, and is therefore applicable to a larger proportion of clinical subjects, than is any task which admits of being performed incorrectly. However, the cost of a form board series is a serious obstacle to its adoption in clinics, while the cost of the color cube outfit is negligible. The writer is not familiar with any other low-priced performance test which compares favorably, for range of applicability, with the color cube series here described.

We are in great need of a performance test battery, each unit of which should be not a detached task but a graded series having a wide range of discriminative capacity. This color cube series can safely be offered as one unit of such a battery.

As a test in itself, it should be used with caution. No claim is made for it as a measure of "intelligence," even on the assumption that intelligence is measurable. Small as is the number of cases at each age, it has been possible to find an eight-year child capable of achieving a superior adult score, and to find several adolescents who were barely able to achieve an eight year score. This high range of variation at each level, as indicated in Table IV, shows it as a test of special aptitude rather than a measure of any capacity which would be found universally among intelligent persons. But if it were invariably used as one of several performance tests, according to the Pintner-Paterson plan (9), the fact that it is a test of special aptitude would strengthen rather than weaken it. Low inter-correlation among the units of a battery is recognized as being advantageous to the battery as a whole.

The test can be recommended for opening an examination, because the compelling interest which it elicits is conducive to a favorable attitude on the part of the subject. This is illustrated by the following note, written recently by an examiner who has no personal interest in the test. "The patient came to the office unwillingly, with an apprehensive and suspicious attitude. She was presented with the color cube test. She started working half-heartedly, grudgingly, and perhaps a little sullenly; but a decided change was noted in her manner as she progressed from

the simpler to the more difficult tasks. She became obviously interested in completing each design. . . . The rest of the examination progressed pleasantly and successfully, with the entire co-operation of the patient." This note refers to a woman of fifty, diagnosed as a case of paranoid condition. She failed to score on the more difficult tasks, and achieved a rating of only eleven; but her response to the challenge of these fascinating tasks carried her through the examination, including the usual language tests. This reaction on the part of a subject is so common as to be observed without surprise by examiners who are accustomed to using the test.

Little can be claimed for the scoring system except that it is a convenient one. The quality of performance, as noted subjectively by the examiner, does not invariably agree with the score. Deliberate and careful procedure, with very few false moves, is more impressive to the observer than is a rapid trial-and-error performance which may quite possibly yield a higher score. For example, the subject may begin by turning all the cubes right side up for this design, instead of turning each one as he places it. Such forethought commands the respect of the observer, but it is penalized in the score. Experiments with trained subjects have established the fact that it is a waste of time to handle each block twice instead of once.

It is of great importance to have a scoring system not wholly dependent upon speed. It might be possible, by means of a cinematographic study, to describe and classify various types of procedure, so that they might be recognized and recorded by name. The significance of any performance test would be vastly increased by an exhaustive laboratory study of all the movements involved.

But a count of gross movements which are noted down as they occur is too inaccurate to add anything of value to the record. The mere recognition of a "move" or "error," however defined, involves such hair-splitting distinctions as to make excessive demands upon personal equation. While the observer is trying to decide whether an uncertain movement should be recorded as a "move" or should be considered too trifling to be counted, a rapidly-working subject can make so many additional

moves as to cause the examiner to lose count. Until the movements can be studied accurately, it seems advisable to base the score upon the variable that can be measured objectively; and to have the examiner's time and attention used for making observations, frankly subjective, concerning the subject's mode of attack.

Incidentally, the test has a value that is wholly independent of numerical results. As a means of observing the subject's ability to take advantage of practice in the performance of similar tasks, it is a more comfortable test than is a practice series involving repetition of the same task. It is much easier to hold the subject's interest in a graded series of increasing difficulty than in a single task which is to be performed repeatedly. Even when there is no stop watch at hand, it is frequently worth while to use the test for observing the subject's ability (or failure) to utilize in the performance of the more difficult tasks what he has learned by performing the earlier and easier ones.

8. SUMMARY

The Kohs Block Design Test has been shortened for general use in the clinic, and has been modified by substituting full-size designs for the smaller designs used by Kohs. The time required for the examination is greatly reduced by the modification.

The materials are compact enough to be carried in a handbag, and their cost is so small as to place the test within reach of clinical workers who have to furnish their own materials.

The test is very attractive to most children and many adults, so it generally commands the spontaneous cooperation of the subject.

It is offered as a test of some special aptitude rather than of mental capacity in general. Its weakness as a measure of intelligence (in so far as intelligence can be measured), is indicated by a very wide range of individual variation at any particular age or mental level.

The test is scored by speed, not because speed is the most important factor involved, but because it is the only variable which can be measured objectively.

The norms are tentative, being based upon about 300 cases.

The usefulness of the test, however, is not wholly dependent upon standardization.

It is a convenient test for examination of a subject whose ability is wholly unknown, because the presentation can be adapted to the individual. The norms are arranged in the form of two overlapping scales, the lower discriminative from 6 to 11 years and the upper from 8 years to the superior adult level. The two scales have enough tasks in common to permit the test to be used as a continuous series, which can be discontinued when the subject reaches his limit.

The test can be presented without use of language, and can therefore be used for immigrants. The norms derived from children do not hold for adults, but the subject's method of procedure is of considerable significance.

The test as modified has been in use ten years, in clinics for children and in state hospitals. It has been found highly serviceable, and in Danvers State Hospital is considered indispensable.

REFERENCES

1. Kohs, S. C. The Block Design Tests. Journal of Experimental Psychology, 1920; pp. 357-376.

2. Kohs, S. C. Intelligence Measurement. Macmillan, 1923.

3. Maxfield, F. N. Design Blocks: A Description of a Simple Clinical Test. Psychological Clinic, 1925; pp. 98-109.

4. Hutt, R. B. W. Standardization of a Color Cube Test. Psychological Clinic, 1925; pp. 77-97.

5. Bronner, Healy, Lowe, Shimberg. A Manual of Individual Mental Tests and Testing. Little, Brown, and Co. 1927.

6. Kent, G. H. Oral Test for Emergency Use in Clinics. Mental Measurements Monographs, No. 9, January, 1932.

7. Kent, G. H. Written Tests for the Clinic. Journal of Genetic Psychology, March, 1934. XLIV, pp. 49-68.

8. Kent, G. H. and Shakow, D. Graded Series of Form Boards. Personnel Journal, August, 1928; VII, pp. 115-120.

9. Pintner, R., and Paterson, D. A Scale of Performance Tests. Appleton, 1917.

PROFILE DRAWINGS OF NORMAL AND SUBNORMAL CHILDREN

EDNA WILLIS McELWEE

New York City Schools

THE present study which makes a comparison between the profile drawings of normal children and subnormal children, was undertaken at the suggestion of Dr. Florence L. Goodenough. In a letter to the writer she made the following statements:

> "There appear to be rather definite qualitative differences, as well as those of a quantitative nature, between the drawings of older defectives and those of the younger normal children. Cyril Burt in particular emphasizes this point. He considers it possible to distinguish between their drawings on the basis of qualitative characteristics alone. While my experience would not lead me to place as much confidence in such a distinction as Burt thinks warranted, it is nevertheless true that a fair percentage of the drawings of older defectives show what I sometimes call the 'primitive profile.' This can best be described as a drawing shown in the profile position but otherwise very primitive in type. Such drawings are rarely made by young children."

The Goodenough drawing test was given to 200 pupils fourteen years of age in the ungraded classes of New York City and also to 200 children in the second and third grades of Public School 208 Brooklyn. All the profile drawings were separated from those made by both groups, the number in each case being rather small. The accompanying figure is an illustration of a primitive profile drawn by one of the subnormal children.

As many as possible of the profile drawings of the younger children were paired with a profile drawing made by an older subnormal child. In each case the mental age of the subnormal child corresponded to the chronological age of the younger child.

Mental ages on the Binet test were available for all the children in the ungraded classes and for some of those in the regular grades. If the Binet test had not been given, a group intelligence test was substituted. With these limitations twelve pairs of drawings were secured.

The fact that the transition period from full-face to profile drawings occurs between eight and ten years of age accounts for the smaller number of profile drawings made by the normal children, most of whom were seven and eight years of age. However, this group was selected because most of the children in the ungraded group had a mental age of seven and eight years. As only a small percentage of the subnormal children made profile drawings, there may also be a similar relationship between profile drawings and mental age.

There are 50 points upon which Dr. Goodenough scores the

TABLE I

A Comparison of the Profile Credit of the Normal and Subnormal Children

Profile Items	Credit	
	Normal	Subnormal
1. Trunk—profile shown by position of arms, pockets, buttons, etc.	8	5
2. Feet—in proportion, both pointing same direction	8	7
3. Legs—upper part of leg in background concealed	2	1
4. Shoulders—indicated by expansion of chest	6	5
5. Arms—attached to shoulders	5	3
6. Nose—differentiated from forehead and upper lip	10	12
7. Mouth—separate modeling of two lips	7	12
8. Nostrils—outline of nose extended inward on upper lip	0	3
9. Ear—detail for aural canal	4	3
10. Eye—glance—pupil toward front	2	3
11. Chin—differentiated from lower lip	4	8
12. Features—coordination of eyes, nose and mouth	0	1
Total	56	63

drawing of a man. From these all were selected which applied to profile drawings as distinct from full-face drawings. Each of the paired drawings was then scored on these twelve items. The number of normal and subnormal children scoring on each of the items is summarized in Table I.

A more detailed comparison of each set of paired drawings, as given in Table II, includes the points in each drawing receiving profile credit, the total score, the mental age of the subnormal child, the chronological age of the normal child, and the IQ of both.

TABLE II

Comparison of Paired Profile Drawings

Pair 1
Normal: Detroit IQ 105; CA 6-10; Total Score 18
Profile Credit: trunk
Subnormal: Binet IQ 48; MA 6-10; Total Score 13
Profile Credit: nose, mouth, chin

Pair 2
Normal: Haggerty IQ 89; CA 7-7; Total Score 22
Profile Credit: trunk, legs, arms, nose, ear
Subnormal: Binet IQ 54; MA 7-6; Total Score 22
Profile Credit: trunk, feet, shoulders, nose, mouth, chin

Pair 3
Normal: Haggerty IQ 101; CA 7-7; Total Score 35
Profile Credit: trunk, feet, shoulders, arms, nose, ear, eye
Subnormal: Binet IQ 56; MA 7-10; Total Score 14
Profile Credit: nose, mouth

Pair 4
Normal: Haggerty IQ 105; CA 8-1; Total Score 16
Profile Credit: nose
Subnormal: Binet IQ 57; MA 8-0; Total Score 20
Profile Credit: trunk, feet, shoulders, nose, mouth

Pair 5
Normal: Binet IQ 125; CA 8-4; Total Score 36
Profile Credit: trunk, feet, shoulders, arms, nose, mouth, ear
Subnormal: Binet IQ 58; MA 8-2; Total Score 23
Profile Credit: nose, mouth, nostrils, chin

Pair 6
Normal: Haggerty IQ 87; CA 8-5; Total Score 28
Profile Credit: trunk, feet, shoulders, arms, nose, mouth
Subnormal: Binet IQ 63; MA 8-9; Total Score 21
Profile Credit: feet, nose, mouth, nostrils, ear

Pair 7
Normal: Haggerty IQ 92; CA 8-9; Total Score 21
Profile Credit: trunk, feet, shoulders, nose, mouth, chin
Subnormal: Binet IQ 64; MA 9-0; Total Score 30
Profile Credit: feet, nose, mouth

Pair 8
Normal: Haggerty IQ 100; CA 8-10; Total Score 29
Profile Credit: trunk, feet, shoulders, nose, mouth, chin
Subnormal: Binet IQ 63; MA 8-10; Total Score 27
Profile Credit: trunk, feet, shoulders, arms, nose, mouth, chin

Pair 9
Normal: Binet IQ 115; CA 9-2; Total Score 19
Profile Credit: feet, eyes
Subnormal: Binet IQ 67; MA 9-5; Total Score 43
Profile Credit: trunk, legs, shoulders, arms, nose, mouth, ear, eyes, chin, features

Pair 10
Normal: Haggerty IQ 95; CA 9-5; Total Score 32
Profile Credit: trunk, feet, legs, shoulders, arms, nose, mouth, chin
Subnormal: Binet IQ 67; MA 9-5; Total Score 26
Profile Credit: feet, nose, mouth, ear, eye, chin

Pair 11
Normal: Haggerty IQ 90; CA 10-0; Total Score 28
Profile Credit: feet, nose, mouth
Subnormal: Binet IQ 74; MA 10-4; Total Score 19
Profile Credit: nose, mouth, eye, chin

Pair 12
Normal: Binet IQ 140; MA 11-4; Total Score 30
Profile Credit: nose, mouth, ear, chin
Subnormal: Binet IQ 78; MA 10-10; Total Score 31
Profile Credit: trunk, feet, shoulders, arms, nose, mouth, nostrils, chin

SUMMARY

An analysis of 12 pairs of profile drawings made by subnormal children and younger normal children, whose chronological ages corresponded with the mental ages of the subnormal children showed:

1. There was a larger number of mature elements in the profiles of the subnormal children. They were more successful than the normal children in representing the features, especially the nose, mouth and chin. They failed more often to show the trunk in profile and to attach the arms at the shoulders.
2. The total score earned by the younger children was 314 and by the subnormal children 289. While the profile drawings of the subnormal children showed more mature elements, they also contained more immature elements, such as absence of trunk, attachment of arms and legs at the neck, a row of buttons to represent clothing, and disproportion of parts. The immature elements more than balanced the mature elements in the total score.

3. In general, the drawings of the subnormal showed more care for detail, fine and delicate shading, and addition of other objects, such as trees, the sun, a telephone pole or a piece of furniture. The most artistic drawing of all was made by one of the subnormal children, but it contained a hand crudely drawn.

ON THE STANDARDIZATION OF THE KUHLMANN REVISION OF THE BINET SCALE

F. H. FINCH

University of Minnesota

FREEMAN[1] and others have called attention to the fact that there exists a definite negative relationship between C.A. and IQ as determined by the Stanford Revision of the Binet Scale. Plotting C. A. against IQ in the original data collected by Terman for the 905 cases upon which the Stanford norms are based revealed an appreciable drop in IQ for those subjects tested at the higher ages. Application of the correlation formula to the same data gave a value of $-.24 \pm .02$, "which is clear evidence that intelligence quotients based upon the Stanford Revision are markedly affected by age." The same paper reports a similar analysis of data from 673 children tested in the University of Chicago Elementary School and from 187 superior children tested by Baldwin at the University of Iowa. The observed correlations between C.A. and IQ were $-.28 \pm .02$ and $-.21 \pm .05$, respectively. Results quoted from other workers using the Stanford Revision agree closely with the above.

There has apparently been no inquiry of a similar nature for the Kuhlmann Revision. Goodenough[2] collected data for a group of children ranging in age from 18 to 54 months which indicate that the norms on this test give results somewhat too high. Her subjects were carefully chosen to afford a representative sample of the population of Minneapolis, appropriate num-

[1]Freeman, Frank N., Holzinger, Karl J., and Mitchell, Blythe C., "The Influence of Environment on the Intelligence, School Achievement, and Conduct of Foster Children." Twenty-seventh Yearbook, National Society for the Study of Education, 1928, pp. 103-217.

[2]Goodenough, Florence L., "The Kuhlmann-Binet Tests for Children of Preschool Age." University of Minnesota Institute of Child Welfare Monographs, No. II, 1928.

bers being taken from each occupational group. The findings were:

Age	2	3	4	Total
N	100	100	100	300
Mean IQ	105.1	104.4	109.4	106.3

For these ages the author suggests "that some modification of the original standards in the direction of more stringent requirements is desirable." (P. 42.)

Mohr and Bartelme[3] report that among 113 children prematurely born who were tested at ages ranging from 8 months to 7 years there existed a correlation of —.34 between C.A. and Kuhlmann IQ. Had these authors corrected ages for prematurity, the relationship would have been even more pronounced than the reported figure indicates.

The data for the present study consist of test records on 553 children who ranged in age from 6 to 15 years when tested with the Kuhlmann Revision.[4] The tests were administered by students who had completed one quarter's work in a course in individual testing, and who had done a limited amount of individual testing under supervision. The records from which the data were taken were collected under the supervision of a competent assistant who checked each test blank with the student tester.

The subjects were from widely scattered sources, many being from public schools in various parts of the city, while a considerable number were relatives or acquaintances of the students doing the testing. While it is impossible to trace out the factors which were effective in determining the selection of the individuals included, there is apparent no influence which would favor the inclusion of bright children of one age and dull children of another. Table I gives the mean intelligence quotients for the various age groups, each group being designated

[3]Mohr, George J., and Bartelme, Phyllis, "Mental and Physical Development of Children Prematurely Born," *American Journal of Diseases of Children*, 1930, 40, 1000-1015.

[4]The test records were made available through the courtesy of Professor John G. Rockwell of the College of Education, Univ. of Minn.

by the beginning point, so that age group 6 includes all who were between 6 and 7 years at the time when they were tested.

TABLE I

Mean IQ for 553 Cases Tested with Kuhlmann Revision Grouped by Chronological Age

Age at test	Mean IQ	S.D.	N
6	106.6	14.2	106
7	103.4	14.9	68
8	100.6	12.6	66
9	100.9	13.8	70
10	102.9	17.1	76
11	102.6	18.0	63
12	99.4	14.4	46
13	103.2	22.7	28
14	102.3	15.2	30
Total	102.8	16.1	553

Despite the statistical objection that may be raised against the use of correlation in a situation of the sort here involved, coefficients have been determined in order to permit comparisons with certain of the results reported by the writers quoted above. For the total group of 553 cases the correlation between C.A. and IQ was found to be $-.079 \pm .03$ (means: C.A., 9.65; IQ, 102.8; sigmas: C.A., 2.41; IQ, 16.06). Since the values for ages 6 and 7 as given in Table I do not agree perfectly with the general trend observed for the higher ages, r was also computed for the 379 cases falling between 8 and 15 years, and was found to be $+.024 \pm .03$ (means: C.A., 10.9; IQ, 101.6; sigmas: C.A., 1.81; IQ, 16.6).

Unless some unrecognized selective factor is operating to influence these results, it seems that for ages 8 to 15 the Kuhlmann Revision is relatively free from an error of standardization known to exist in the Stanford Revision and which is also probably present in the norms on the Kuhlmann Revision for earlier ages. While the results from the present records for age 6 agree in a measure with those previously reported, the possibility that some unknown selective factor exists is so great that no scheme for correction of results at this age is offered. Such a correction requires a comprehensive study based upon data similar to those collected by Goodenough but covering the total age range to which it is to be applied.

NEWS AND NOTES

The University of Chicago has recently set up an organization to provide for specialization in Child Development at the level of the Ph.D. The plan furnishes a general background of instruction in the basic sciences contributing to an understanding of Child Development, such as psychology and biology, and of the social institutions which exert the chief influence on the child's development, such as the school and the family. It also provides the opportunity for specialization and research in any aspect of development, such as physical growth, mental growth, behavior, nutrition, education, the influence of social factors, heredity and environment, etc.

For the present the programs of instruction and research are organized and supervised by the Department of Education in cooperation with the University Committee on Child Development. This Department accepts programs of work, planned and carried out in each case with the supervision of a special committee, which include appropriate courses in any department of the University. Departmental requirements have been modified in the case of students in Child Development to make possible the inclusion of this work from other departments. The committee which supervises the student's program includes a representative of the Department of Education, the Committee on Child Development, and the department in which the student's research lies.

The University of Chicago has for some time been provided with a considerable list of courses of instruction, facilities for research and laboratory schools, including a school for problem children. It now has the machinery by which these facilities can be more effectively utilized by the student who wishes to specialize in Child Development than was formerly the case.

Teachers of vocational agriculture in rural high schools of the United States are devoting increasing attention to instruction in farm problem analysis, according to reports to the United States Office of Education. Before teachers attempt to give actual instruction in the growing of farm crops, they point out to the students certain decisions they should make, factors they should consider, and the kind of information they need in preparing to grow each crop. One of the latest education bulletins issued by the Federal Office is entitled "Analysis of the Management of a Cotton-Growing Enterprise."

More than half the public school construction now going on in the United States is being financed through PWA, recently announced Administrator Harold L. Ickes. With the beginning of the new scholastic year more than $210,000,000 have been added to the capital outlay for school building construction as a direct result of the $3,700,000,000 PWA program.

BOOK REVIEWS

Walker, Helen M. *Mathematics Essential for Elementary Statistics,* Henry Holt and Co., New York, 1934, xiii-246.

Students taking courses in elementary statistics have needed such a book as this one. The topics apparently have been selected by an empirical study of the difficulties of such students. The topics included such as the following: Significant figures, fractions, decimal points, reading tables of squares and square roots, computational short cuts, simple equations, graphs, and logarithms, factoring and summation, and others. The fact that it is a self teaching book is commendable. Each chapter is prefaced by a first test, by which the student can judge whether or not he needs to study that chapter. The matters of selection and interpretation are not within the scope of this book.

The book is full of examples which are clear, and has practice exercises adequate both in variety and quantity for students to learn the topics being presented. Answers to the tests and exercises are given at the back of the book. At the end of most chapters there is a second test to enable a student to find out whether or not he has learned the principles presented in that chapter.

The book should be useful not only for the students of elementary statistics but also in training statistical clerks.

This book might well be followed by a volume on the more advanced mathematics essential to a background in statistics.

HAROLD A. EDGERTON
Ohio State University

William Lewin. *Photoplay Appreciation in American High Schools:* D. Appleton-Century Company, New York. $1.00.

The nation-wide interest recently aroused in the subject of motion pictures and their influence, particularly upon youth, makes timely the publication this month (September) of a monograph by William Lewin, *Photoplay Appreciation in American High Schools,* sponsored by the National Council of Teachers of English.

Photoplays tend to have a more constructive influence upon the attitude and conduct of children who receive classroom instruction in motion picture criticism, the author concludes after a study of the film reactions reported by 1851 high school pupils. Where class discussions of pictures had been held, a higher percentage was able to mention instances of specific influence, usually in the direction of higher ideals.

Educators, clergymen, and parents have long discussed with grave concern the "motion picture problem," but not until two and a half years ago was any national, systematic effort made to test the possibilities of solution through school instruction. At that time, the National Council of Teachers of English formed a Photoplay Appreciation Committee, with Dr. Lewin as chairman, and began a pioneer experiment to determine whether the motion picture standards and tastes of adolescents might be improved through the medium of the English class. The methods and findings of the experiment, which has now assumed the proportions of a widespread movement looking toward the introduction of photoplay appreciation study in high schools, are comprised in the book just published.

Children see approximately fifty photoplays a year, yet three-quarters of them have never written a theme or prepared a talk about a film, the committee discovered. This suggested an educational opportunity ignored.

Teachers of high school English classes scattered throughout the country agreed to test the results of class criticism of pictures. They set up two groups of pupils—sixty-eight in all—evenly matched in age, grade, intelligence, and home background. Both groups went in a body to see the same series of photoplays, but only one received instruction. By means of statistical procedures, it was found that the experimental pupils showed a superiority of 41.4 per cent. over control pupils in judging plays seen.

Pupils under guidance, it is reported, soon formed the habit of seeking the teacher's advice before seeing a picture, something unheard of before. The critical discussions and technical analyses tended to increase enjoyment of pictures and to develop greater appreciation of the portrayal of such ideals as honesty, bravery, devotion and self-sacrifice.

Both teachers and pupils, according to quoted comments, enthusiastically approved the teaching innovation, and Dr. Lewin adds, "With the assistance of the photoplay, the work of teachers of literature becomes more interesting to the pupil, and the necessity of compelling attention through discipline virtually disappears."

Outlines of units of instruction in photoplay appreciation for junior and senior high schools, suggestions for activities of a photoplay club, and several stenographic reports of class discussions are included in the monograph, which is of especial interest to high school teachers and administrators.

National Council of Teachers of English

NEW BOOKS AND PAMPHLETS RECEIVED

Books and pamphlets for review should be sent to James P. Porter, Editor, Journal of Applied Psychology, Ohio University, Athens, Ohio.

Can Attitudes Be Taught? Arthur Lichtenstein. Johns Hopkins Studies in Education, No. 21. Johns Hopkins Press, Baltimore, 1934. $1.25. 89 pp.

Child Psychology. George D. Stoddard and Beth L. Wellman. Macmillan Company, New York, 1934. $2.50. 419 pp.

General Experimental Psychology. Arthur Gilbert Bills. Longmans, Green & Company, New York, 1934. $4.00. 620 pp.

A Guide to Reading in Aesthetics and Theory of Poetry. George N. Belknap. University of Oregon Publications, Vol. IV, No. 9. University of Oregon, Eugene, Oregon. 91 pp.

The Horizontal Organization of Secondary Education. U. S. Department of the Interior, Bulletin No. 17, 1932. Government Printing Office, Washington, D. C., 1934. 20 cents. 273 pp.

Mental Hygiene and Education. Mandel Sherman. Longmans, Green & Company, New York, 1934. $2.25. 295 pp.

National Survey of Education. Bulletin No. 17, 1923. U. S. Department of the Interior, Office of Education, Washington, D. C., 1934. 231 pp.

Predisposition Aux Accidents. Eug. Schreider. Conservatoire National des Arts et Metiers. 292 Rue Saint-Martin, Paris. 60 pp.

The Social Science Research Council—Decennial Report 1923-33. New York City, 132 pp.

Directory of American Psychological Periodicals

(Continued)

Comparative Psychology Monographs—Baltimore: Johns Hopkins University Press. Subscription $5. 400 pages per volume. Knight Dunlap, Managing Editor. Published without fixed dates, each number a single research.

Genetic Psychology Monographs—Worcester, Mass.: Clark University Press. Subscription $7.00 per volume. Two volumes per year, 600 pages each. Edited by Carl Murchison. Monthly. Each number one complete research. Child behavior, animal behavior and comparative psychology. Founded 1925.

Psychological Abstracts—Eno Hall, Princeton, N. J. American Psychological Association. Subscription $6.00. 700 pages ann. Edited by W. S. Hunter. Monthly. Abstracts of psychological literature. Founded 1927.

Pedagogical Seminary and Journal of Genetic Psychology—Worcester, Mass.: Clark University Press. Subscription $14.00 per annum; $7.00 per volume. 1,000 pages annually (2 volumes). Edited by Carl Murchison and an international cooperating board. Quarterly. Child behavior, animal behavior, and comparative psychology. Founded 1891.

Journal of Social Psychology—Worcester, Mass.: Clark University Press. Subscription $7.00. 500-600 pages annually. Edited by John Dewey and Carl Murchison. Quarterly. Political, racial, and differential psychology. Founded 1929.

Personnel Journal—Baltimore: The Williams & Wilkins Co. Subscription $5.00. 500 pages, approximately. Edited by Walter V. Bingham. Bi-Monthly.

Mental Measurement Monographs—Baltimore: The Williams & Wilkins Company. Subscription $5. 400 pages annually. Edited by Knight Dunlap and Buford Johnson. Published without fixed dates, each number a single research. Founded 1925.

Studies in Psychology and Psychiatry—Baltimore: The Williams & Wilkins Company. Subscription $5.00. 400 pages annually. Edited by Edward A. Pace. Published without fixed dates, each number a single research. Founded 1926.

Journal of Philosophy—New York: 515 West 116th Street. Subscription $4.00. 728 pages per volume. Founded 1904. Bi-weekly. Edited by F. J. E. Woodbridge, Wendell T. Bush, and H. W. Schneider.

Child Development—Baltimore, Md.: Williams & Wilkins Company. Subscription $5.00. 400 pages annually. Edited by Buford Johnson. Quarterly. Founded 1930.

Character and Personality—Durham, N. C.: Duke University Press. Subscription $2.00 per annum. Quarterly. Edited by Robert Saudek (London) and Ernest Seeman (Durham, N. C.). Founded 1932.

Directory of American Psychological Periodicals

American Journal of Psychology—Ithaca, N. Y.: Cornell University. Subscription $6.50. 624 pages annually. Edited by M. F. Washburn, K. M. Dallenbach, M. Bentley, E. G. Boring. Quarterly. General and experimental psychology. Founded 1887.

Journal of General Psychology—Worcester, Mass.: Clark University Press. Subscription $14.00 per annum; $7.00 per volume. 1,000 pages annually (2 volumes). Edited by Carl Murchison and an international cooperating board. Quarterly. Experimental, theoretical, clinical, and historical psychology. Founded 1927.

Psychological Review—Princeton, N. J.: Psychological Review Company. Subscription $5.50. 540 pages annually. Bi-monthly. General Psychology. Edited by Howard C. Warren. Founded 1894.

Psychological Monograph—Princeton, N. J.: Psychological Review Company. Subscription $6.00 per vol. 500 pp. Ed. by Raymond Dodge. Published without fixed dates, each issue one or more researches. Founded 1895.

Psychological Index—Princeton, N. J.: Psychological Review Company. Subscription $4.00. 300-400 pp. Edited by W. S. Hunter. An annual bibliography of psychological literature. Founded 1895.

Psychological Bulletin—Princeton, N. J.: Psychological Review Company. Subscription $6.00. 720 pages annually. Psychological literature. Monthly. Edited by Edward S. Robinson. Founded 1904.

Archives of Psychology—New York: Columbia University. Subscription $6. 500 pp. ann. Ed. by R. S. Woodworth. Published without fixed dates, each number a single experimental study. Founded 1906.

Journal of Abnormal and Social Psychology—Eno Hall, Princeton, N. J. American Psychological Association. Subscription $5. Foreign, $5.25. Editor, Henry T. Moore. Quarterly. 448 pages annually. Abnormal and Social. Founded 1906.

Psychological Clinic—Philadelphia: Psychological Clinic Press. Subscription $3.00. 288 pages. Ed. by Lightner Witmer. Without fixed dates (9 numbers). Orthogenics, psychology, hygiene. Founded 1907.

Psychoanalytic Review—Washington, D. C.: 3617 10th St., N. W. Subscription $6. 500 pages annually. Psychoanalysis. Quarterly. Ed. by W. A. White and S. E. Jelliffe. Founded 1913.

Journal of Experimental Psychology—Princeton, N. J. Psychological Review Company. 500 pages annually. Experimental. Subscription $6.00. Bi-monthly. Edited by S. W. Fernberger. Founded 1916.

Journal of Applied Psychology—Crawfordsville, Ind.: R. R. Donnelley & Sons Co. Subscription $5.00. 600 pages annually. Bi-monthly. Edited by James P. Porter, Ohio University, Athens, Ohio. Founded 1917.

Journal of Comparative Psychology—Baltimore: The Williams & Wilkins Co. Subscription $5.00 per volume of 450 pages. Three volumes every two years. Edited by Knight Dunlap and Robert M. Yerkes. Founded 1921.

(Continued on Third Cover)